

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 5.60748 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-1

Perfect score: 28

Sequence: 1 RGRPQ 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir.78.\*

2: Pir.2.\*

3: Pir.3.\*

4: Pir.4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	79	2 A44843	EMZp30-47 antigen
2	28	100.0	137	1 JC4877	Phospholipase A2 h
3	28	100.0	138	2 T12631	DRS12 protein - co
4	28	100.0	141	2 T00328	hypothetical prote
5	28	100.0	141	2 T46015	hypothetical prote
6	28	100.0	153	2 A34132	vasotocin / neurop
7	28	100.0	165	2 T25378	hypothetical prote
8	28	100.0	166	1 P1HUSC	salivary proline-r
9	28	100.0	166	2 B25372	salivary proline-r
10	28	100.0	171	2 A27307	proline-rich phosph
11	28	100.0	177	2 G75285	hypothetical prote
12	28	100.0	179	2 B72584	hypothetical prote
13	28	100.0	197	2 S26438	hypothetical prote
14	28	100.0	217	2 S26452	hypothetical prote
15	28	100.0	225	2 G75448	conserved hypothet
16	28	100.0	276	2 F53161	alpha 2-adrenergic
17	28	100.0	289	2 B72625	probable acetylglu
18	28	100.0	301	2 H75493	arginase - Deinoco
19	28	100.0	301	2 B75509	conserved hypothet
20	28	100.0	307	1 TQEC93	transposase - Esch
21	28	100.0	320	2 G75284	probable polyprotein
22	28	100.0	369	2 S21471	genome polyprotein
23	28	100.0	382	2 S71281	cytochrome c-type
24	28	100.0	445	2 T42203	serotonin receptor
25	28	100.0	450	2 A37223	alpha-2B-adrenergic
26	28	100.0	469	1 NMIV27	exo-alpha-sialidas
27	28	100.0	482	2 F69357	conserved hypothet
28	28	100.0	520	2 T33146	hypothetical prote
29	28	100.0	577	2 S77715	cytochrome c-type

cytochrome c-type  
hypothetical prote  
ZP3 receptor presu  
collagen alpha 2(X  
probable transcrip  
(p)pppp 3-pyropho  
hypothetical prote  
probable phosphor  
hypothetical prote  
proline dehydrogen  
hypothetical prote  
probable phosphor  
hypothetical prote  
ribonucleoside-dip  
probable transposo  
neural cell adhesi

#### ALIGNMENTS

##### RESULT 1

A44843  
EMZp30-47 antigen - Eimeria acervulina (fragment)  
C:Species: Eimeria acervulina  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 23-Jun-1993  
C:Accession: A44843  
R:Jenkins, M.C.; Lillehoj, H.S.; Barta, J.R.; Danforth, H.D.; Strohllein, D.A.  
Exp. Parasitol. 70, 353-362, 1990  
A:Title: Eimeria acervulina: cloning of a cDNA encoding an immunogenic region of seve  
A:Reference number: A44843; MUID:90184343; PMID:1690144  
A:Accession: A44843  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-79 <JEN>  
A:Cross-references: GB:M37843

Query Match 100.0%; Score 28; DB 2; Length 79;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5

DB 6 RGRPQ 10

##### RESULT 2

JC4877  
phospholipase A2 homolog 3 precursor - himenhabu  
C:Species: Trimeresurus okinavensis (himenhabu)  
C:Date: 15-Aug-1996 #sequence\_revision 14-Feb-1997 #text\_change 11-Jun-1999  
C:Accession: JC4877  
R:Nobuhisa, I.; Nakashima, K.; Deshimaru, M.; Ogawa, T.; Shimohigashi, Y.; Fukumaki, Y  
Gene 172, 267-272, 1996  
A:Title: Accelerated evolution of Trimeresurus okinavensis venom gland phospholipase  
A:Reference number: JC4874; MUID:96269416; PMID:8682315  
A:Contents: venom gland  
A:Accession: JC4877  
A:Molecule type: mRNA  
A:Residues: 1-137 <NO>  
A:Cross-references: DBJ:D49389; MID:g1469806; PIDN:BA08384.1; PID:g1469807  
C:Comment: This protein probably does not have phospholipase A2 activity.  
C:Superfamily: phospholipase A2  
C:Keywords: venom

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-137/Product: phospholipase A2 homolog 3 #status predicted <MAT>

F:42-131,44-60,59-111,65-137,66-104,73-97,91-102/disulfide bonds: #status predicted

Query Match 100.0%; Score 28; DB 2; Length 137;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5

Db 49 RGRPQ 53  
|||||

## RESULT 3

T2631  
DS12 protein - common sunflower (fragment)  
C:Species: Helianthus annuus (common sunflower)  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 13-Aug-1999  
C:Accession: T12631  
R:Steinmetz, A.  
Submitted to the EMBL Data Library, March 1997  
A:Reference number: Z17560  
A:Accession: T12631  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-138 <ST>  
A:Cross-references: EMBL:Y11903; NID:e1041628; PID:e308647  
C:Genetics:  
A:Gene: DS12

Query Match 100.0%; Score 28; DB 2; Length 138;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||||  
Db 80 RGRPQ 84

## RESULT 4

T00928  
Hypothetical protein At2g42190 [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein T24P15.10  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 23-Mar-2001  
C:Accession: T00928; H84850  
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
Submitted to the EMBL Data Library, December 1997  
A:Description: Arabidopsis thaliana chromosome II BAC T24P15 genomic sequence.  
A:Reference number: Z14212  
A:Accession: T00928  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-141 <ROU>  
A:Cross-references: EMBL:AC002561; NID:g2673901; PID:g2673910  
A:Experimental source: cultivar Columbia  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: H84420; MUID:20083487; PMID:10617197  
A:Accession: H84850  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-141 <STO>  
A:Cross-references: GB:AB002093; NID:g2673910; PID:AA88644.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: T24P15.10; At2g42190  
A:Map position: 2  
C:Superfamily: Arabidopsis thaliana hypothetical protein T10K37.140

Query Match 100.0%; Score 28; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||||  
Db 33 RGRPQ 37

## RESULT 5

T46015  
Hypothetical protein T10K17.140 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 28-Jul-2000  
C:Accession: T46015  
R:Benes, V.; Wurnbach, E.; Drzonek, H.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.  
Submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23019  
A:Accession: T46015  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-141 <BEN>  
A:Cross-references: EMBL:AL132977  
A:Experimental source: cultivar Columbia; BAC clone T10K17  
C:Genetics:  
A:Map position: 3  
A:Note: T10K17.140  
C:Superfamily: Arabidopsis thaliana hypothetical protein T10K17.140

Query Match 100.0%; Score 28; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||||  
Db 33 RGRPQ 37

## RESULT 6

A34132  
Vasotocin / neurophysin 1 precursor - chum salmon  
N:Contains: neurophysin; vasotocin  
C:Species: Oncorhynchus keta (chum salmon)  
C:Date: 18-May-1990 #sequence\_revision 18-May-1990 #text\_change 16-Jul-1999  
C:Accession: A34132; JCI491  
R:Heierhorst, J.; Mahlmann, S.; Morley, S.D.; Coe, I.R.; Sherwood, N.M.; Richter, D.  
FEBS Lett. 260, 301-304, 1990  
A:Title: Molecular cloning of two distinct vasotocin precursor cDNAs from chum salmon  
A:Reference number: A34132; MUID:90127450; PMID:2298304  
A:Accession: A34132  
A:Molecule type: mRNA  
A:Residues: 1-153 <HEI>  
A:Cross-references: GB:X17327; NID:g64160; PID:CAA35205.1; PID:g64161  
R:Urano, A.

Submitted to JIPID, April 1992  
A:Description: Cloning and sequence analyses of cDNAs encoding vasotocin and isotocin  
A:Reference number: JCI489

A:Accession: JCI491  
A:Molecule type: mRNA  
A:Residues: 1-6,'Q',8-153 <HYO>  
C:Superfamily: oxytocin-neurophysin  
C:Keywords: hormone; neuropeptide  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-29/Product: vasotocin #status predicted <MAY>  
F:33-153/Product: neurophysin #status predicted <NEU>

Query Match 100.0%; Score 28; DB 2; Length 153;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||||  
Db 149 RGRPQ 153

## RESULT 7

T25378  
Hypothetical protein T27F2.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25378  
R:Lennard, N.



submitted to the EMBL Data Library, June 1996

A:Reference number: Z20025  
A:Accession: T25378  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-165 <WIL>  
A:Cross-references: EMBL:Z74045; NID:e1062212; PIDN:CAA98551.1; GSPDB:GN00023; CESP:T27F  
A:Experimental source: clone T27F2  
C:Genetics:  
A:Gene: CESP:T27F2.4  
A:Map position: 5  
A:Introns: 63/3

Query Match 100.0%; Score 28; DB 2; Length 165;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
DB 79 RGRPQ 83

# RESULT 8

FIHUSC  
salivary proline-rich phosphoprotein precursor PRH2 (validated) - human  
N:Alternate names: salivary acidic proline-rich protein PRH2  
N:Contains: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotein  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1981; Revision 12-Apr-1996; #text change 08-Dec-2000  
A:Accession: A25372; A19803; B57868; A92277; A92254; A94725; A91954; S02564; S02563; JPO  
R:Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
J. Biol. Chem. 260, 1123-1130, 1985

A:Title: Differential RNA splicing and post-translational cleavages in the human salivary  
A:Reference number: A92492; MUID:85289325; PMID:2993301

A:Accession: A25372  
A:Molecule type: mRNA

A:Residues: 1-166 <MB>  
A:Cross-references: GB:K03202; NID:g190481; PIDN:AAA60183.1; PID:g190482

R:Schlesinger, D.H.; Hay, D.I.  
Int. J. Pept. Protein Res. 17, 34-41, 1981

A:Title: Primary structure of the active tryptic fragments of human and monkey salivary  
A:Reference number: A91757; MUID:81191179; PMID:7228490

A:Accession: A19803

A:Molecule type: protein

A:Residues: 17-46 <SCH>

R:Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A:Title: Structures of two HaeIII-type genes in the human salivary proline-rich protein

A:Reference number: A57868; MUID:86196106; PMID:3009472

A:Accession: B57868

A:Molecule type: DNA

A:Residues: 1-166 <KIM>

A:Cross-references: GB:M13058; NID:g190513; PIDN:AAA98808.1; PID:g190514

R:Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A:Title: The primary structure of a salivary calcium-binding proline-rich phosphoprotein  
A:Reference number: A92277; MUID:80204368; PMID:7380845

A:Contents: protein C

A:Accession: A92277

A:Molecule type: protein

A:Residues: 17-19, 'N', 21-166 <WON>

A:Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite forma

R:Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A:Title: The complete primary structure of a proline-rich phosphoprotein from human sali  
A:Reference number: A92254; MUID:79173237; PMID:438215

A:Contents: protein A

A:Accession: A92254

A:Molecule type: protein

A:Residues: 17-19, 'N', 21-122 <W02>

R:Schlesinger, D.H.; Hay, D.I.

In Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gross

A:Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A:Reference number: A94425

A:Accession: A94425

A:Molecule type: protein

A:Residues: 17-122 <SC2>

A:Note: the authors call this protein PRP-4

R:Isemura, S.; Saichou, E.; Sanada, K.

J. Biochem. 87, 1071-1077, 1980

A:Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relat

A:Reference number: A91954; MUID:80227634; PMID:7390979

A:Contents: peptide P-C

A:Accession: A91954

A:Molecule type: protein

A:Residues: 123-166 <ISE>

R:Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluc

Biochem. J. 255, 15-21, 1988

A:Title: The primary structures of six human salivary acidic proline-rich proteins (PRP

A:Reference number: S02562; MUID:89061650; PMID:3196309

A:Accession: S02564

A:Molecule type: protein

A:Residues: 17-166 <HAY>

A:Accession: S02563

A:Molecule type: protein

A:Residues: 47-71 <HA2>

R:Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 27, 373-379, 1986

A:Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibi

A:Reference number: JPO106; MUID:86222916; PMID:3710693

A:Accession: JPO106

A:Molecule type: protein

A:Residues: 17-161, 'Q', 163-166 <SC3>

A:Experimental source: parotid gland

R:Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.

Biochemistry 30, 3351-3356, 1991

A:Title: Basic proline-rich proteins from human parotid saliva: relationships of the co

A:Reference number: A38355; MUID:91190884; PMID:1849422

A:Accession: G38355

A:Status: preliminary

A:Molecule type: protein

A:Residues: 123-166 <KAY>

R:Robinson, R.; Kaufman, D.L.; Waye, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.

Biochem. J. 283, 491-503, 1989

A:Title: Primary structure and possible origin of the non-glycosylated basic proline-ri

A:Reference number: S06153; MUID:90088384; PMID:2688632

A:Accession: S06153

A:Molecule type: protein

A:Residues: 123-166 <ROB>

C:Comment: The proposed biological functions are a highly potent inhibitor of crystal g

A:Gene: GDB:PRH2

A:Cross-references: GDB:119516; OMIM:168790

A:Map position: 12p13.2-12p13.2

A:Introns: 22/1; 34/1

C:Superfamily: proline-rich protein

C:Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-166/Product: protein C #status experimental <PRC>

F:17-122/Product: protein A #status experimental <PRA>

F:17-46/Region: apatitic mineral binding

F:47-71/Product: PRP-3 #status experimental <PRP3>

F:123-166/Product: peptide P-C #status experimental <PPC>

F:17/Modified site: pyroglutamate carboxylic acid (Gln) (in mature form) #status experime

F:24/38/Binding site: phosphatase (Ser) (covalent) #status experimental

Query Match 100.0%; Score 28; DB 1; Length 166;

Best Local Similarity 100.0%; Pred. No. 59;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5

DB 122 RGRPQ 126

RESULT 9

```

B25372
salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human
C:Species: Homo sapiens (man)
C>Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 20-Aug-1999
C:Accession: B25372; A57868; S02562; G38355; S06153; B27307
R:Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.
J. Biol. Chem. 260, 11123-11130, 1985
A:Title: Differential RNA splicing and post-translational cleavages in the human salivary
A:Reference number: A92492; MUID:85289325; PMID:2993301
A:Accession: B25372
A:Molecule type: mRNA
A:Residues: 1-166 <MAE>
A:Cross-references: GB:K03203; NID:G190483; PIDN:AAA60184.1; PID:G190484
R:Kim, H.S.; Maeda, N.
J. Biol. Chem. 261, 6712-6718, 1986
A:Title: Structures of two HaeIII-type genes in the human salivary proline-rich protein
A:Reference number: A57868; MUID:86196106; PMID:3009472
A:Accession: A57868
A:Molecule type: DNA
A:Residues: 1-166 <KIM>
A:Cross-references: GB:M13057; NID:G190511; PIDN:AAA98807.1; PID:G190512
R:Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck
Biochem. J. 255, 15-21, 1988
A:Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-
A:Reference number: S02562; MUID:89061650; PMID:3196309
A:Accession: S02562
A:Molecule type: Protein
A:Residues: 47-71 <HAY>
R:Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
Biochemistry 30, 3351-3356, 1991
A:Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov
A:Reference number: A38355; MUID:91190884; PMID:1849422
A:Accession: G38355
A:Molecule type: Protein
A:Residues: 123-166 <KAU>
R:Robinson, R.; Kauffman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.
Biochem. J. 263, 497-503, 1989
A:Title: Primary structure and possible origin of the non-glycosylated basic proline-ric
A:Reference number: S06153; MUID:90088384; PMID:2688632
A:Accession: S06153
A:Molecule type: Protein
A:Residues: 123-166 <ROB>
R:Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.
Am. J. Hum. Genet. 41, 1035-1047, 1987
A:Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich pro
A:Reference number: A27307; MUID:88074309; PMID:3687941
A:Contents: allele pa
A:Accession: B27307
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown
A:Residues: 17-41, 'L', 43-118, 'C', 120-166 <AZE>
A:Cross-references: EMBL:K03203
C:Genetics:
A:Gene: GDB:PRH1
A:Cross-references: GDB:119515; OMIM:168730
A:Map position: 12p13.2-12p13.2
A:Introns: 22/1, 34/1
C:Superfamily: proline-rich protein
C:Keywords: phosphoprotein; saliva; tandem repeat

Query Match 100.0%; Score 28; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5
DB 122 RGRPQ 126

RESULT 10
A27307
N:Alternate names: salivary acidic proline-rich protein

C:Species: Homo sapiens (man)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 29-Aug-1997
C:Accession: A27307
R:Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.
Am. J. Hum. Genet. 41, 1035-1047, 1987
A:Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich p
A:Reference number: A27307; MUID:88074309; PMID:3687941
A:Accession: A27307
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-171 <AZE>
A:Cross-references: EMBL:K03203
C:Genetics:
A:Gene: GDB:PRH1
A:Cross-references: GDB:119515; OMIM:168730
A:Map position: 12p13.2-12p13.2
C:Superfamily: proline-rich protein
C:Keywords: phosphoprotein

Query Match 100.0%; Score 28; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5
DB 127 RGRPQ 131

RESULT 11
G75285
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: G75285
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, B.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: G75285
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <WHI>
A:Cross-references: GB:AE002065; NID:G6460149; PIDN:AAF11894.1; PID:G64601
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2344
A:Map position: 1

Query Match 100.0%; Score 28; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5
DB 141 RGRPQ 145

RESULT 12
B72584
hypothetical protein APB1950 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: B72584
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Taka
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: B72584
A:Status: preliminary
A:Molecule type: DNA

```

A;Residues: 1-179 <RAW>  
 A;Cross-references: DDBJ:AF000062; NID:g5105244; PIDN:BA80959.1; PID:g5105646  
 A;Experimental source: strain K1  
 C;Genetics:  
 A;Gene: APE1950

Query Match 100.0%; Score 28; DB 2; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
 |||||  
 Db 81 RGRPQ 85

## RESULT 13

S26438  
 hypothetical protein 3 - Methanobacterium thermoformicicum plasmid pFZ1  
 C;Species: Methanobacterium thermoformicicum  
 C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 22-Oct-1999  
 C;Accession: S30303; S26438  
 R;Noelling, J.; van Eeden, F.J.M.; Eggen, R.I.L.; de Vos, W.M.  
 Nucleic Acids Res. 20, 6501-6507, 1992  
 A;Title: Modular organization of related Archaeal plasmids encoding different restriction  
 A;Reference number: S30302; MUID:93126090; PMID:1336177  
 A;Accession: S30303  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-197 <NOE>  
 A;Cross-references: EMBL:X68366; NID:g44632; PIDN:CAA48427.1; PID:g44634  
 C;Genetics:  
 A;Genome: plasmid

Query Match 100.0%; Score 28; DB 2; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
 |||||  
 Db 118 RGRPQ 122

## RESULT 14

S26452  
 hypothetical protein 3' - Methanobacterium thermoformicicum plasmid pFZ1  
 C;Species: Methanobacterium thermoformicicum  
 C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 22-Oct-1999  
 C;Accession: S30318; S26452  
 R;Noelling, J.; van Eeden, F.J.M.; Eggen, R.I.L.; de Vos, W.M.  
 Nucleic Acids Res. 20, 6501-6507, 1992  
 A;Title: Modular organization of related Archaeal plasmids encoding different restriction  
 A;Reference number: S30302; MUID:93126090; PMID:1336177  
 A;Accession: S30318  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-217 <NOE>  
 A;Cross-references: EMBL:X68367; NID:g44646; PIDN:CAA48442.1; PID:g44650  
 C;Genetics:  
 A;Genome: plasmid

Query Match 100.0%; Score 28; DB 2; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
 |||||  
 Db 138 RGRPQ 142

## RESULT 15

G75448  
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
 C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C;Accession: G75448  
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vanatkevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.W.  
 Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: G75448  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-225 <WHI>  
 A;Cross-references: GB:AE001952; GB:AE000513; NID:g6458725; PIDN:AAF10575.1; PID:g6458  
 A;Experimental source: strain R1  
 C;Genetics:  
 A;Gene: DR0999  
 A;Map position: 1

Query Match 100.0%; Score 28; DB 2; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
 |||||  
 Db 78 RGRPQ 82

Search completed: April 6, 2004, 16:16:49  
 Job time : 6.60748 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 3.2443 seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-1  
Perfect score: 28  
Sequence: 1 RGRPQ 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141691 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	79	1	CMC1_EIMAC
2	28	100.0	124	1	CES4_HUMAN
3	28	100.0	137	1	PA23_TRIOK
4	28	100.0	153	1	NEU3_ONCKE
5	28	100.0	166	1	PRPC_HUMAN
6	28	100.0	197	1	YP23_METTF
7	28	100.0	217	1	YP23_METTF
8	28	100.0	307	1	TRA2_ECOLI
9	28	100.0	358	1	RLUD_RALSO
10	28	100.0	360	1	PADH_ANYME
11	28	100.0	384	1	A2AB_ELEMA
12	28	100.0	387	1	A2AB_MACPR
13	28	100.0	388	1	A2AB_ORIAT
14	28	100.0	389	1	A2AB_HORSE
15	28	100.0	389	1	A2AB_PROHA
16	28	100.0	390	1	A2AB_DUGDU
17	28	100.0	391	1	A2AB_ERIEU
18	28	100.0	404	1	CT99_HUMAN
19	28	100.0	450	1	A2AB_HUMAN
20	28	100.0	451	1	NRAM_IACKQ
21	28	100.0	451	1	NRAM_IACKR
22	28	100.0	469	1	NRAM_IARIS
23	28	100.0	577	1	CCBS_ORNBE
24	28	100.0	579	1	CCBS_DAUCA
25	28	100.0	977	1	BAB1_DROME
26	28	100.0	1067	1	BAB2_DROME
27	28	100.0	1137	1	RIR1_HSV11
28	28	100.0	1259	1	CAML_RAT
29	28	100.0	1260	1	CAML_MOUSE
30	28	100.0	1305	1	FTSK_YERPE
31	28	100.0	1329	1	FTSK_ECOLI
32	28	100.0	1342	1	FTSK_ECO57
33	28	100.0	1342	1	FTSK_SHIFL

34 28 100.0 1343 1 FTSK\_SALTI  
35 28 100.0 1347 1 FTSK\_ECOL6  
36 28 100.0 1351 1 FTSK\_SALTY  
37 28 100.0 1556 1 BABA\_HUMAN  
38 28 100.0 1736 1 CA2B\_MOUSE  
39 25 89.3 95 1 Y10K\_HPBVY  
40 25 89.3 116 1 V10K\_HPBVY  
41 25 89.3 171 1 V19\_TBSV8  
42 25 89.3 172 1 V19\_AMCV  
43 25 89.3 172 1 V19\_TESVA  
44 25 89.3 288 1 HOL\_PIG  
45 25 89.3 289 1 HOL\_MOUSE

## ALIGNMENTS

RESULT 1  
CMC1\_EIMAC STANDARD; PRT; 79 AA.  
ID CMC1\_EIMAC  
AC P21959;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE EAWZF30-47 protein (Fragment).  
GN CMC17.  
OS Eimeria acervulina.  
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
OC Eimeria.  
OX NCBI\_TaxID=5801;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90184343; PubMed=1690144;  
RA Jenkins M.C., Lillehoj H.S., Barta J.R., Danforth H.D.,  
RA Stroblein D.A.;  
RT "Eimeria acervulina: cloning of a cDNA encoding an immunogenic region  
of several related merozoite surface and rhoptry proteins.";  
RL Exp. Parasitol. 70:353-362(1990).  
CC -!- SUBCELLULAR LOCATION: Surface membrane and internal rhoptries.  
CC -!- DEVELOPMENTAL STAGE: Merozoite.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M37843; AAA62796.1; -.  
KW Antigen; Membrane; Merozoite.

FT NON\_TER 1  
FT NON\_TER 79

SQ SEQUENCE 79 AA; 8703 MW; 0B2E6CDE565FB4330 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 79;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 5; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 RGRPQ 5

Db 6 RGRPQ 10

## RESULT 2

CES4\_HUMAN STANDARD; PRT; 124 AA.  
ID CES4\_HUMAN

AC Q9BXQ7;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cat eye syndrome critical region protein 4 (Fragment).

GN CECR4.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
FN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21275466; PubMed=11381032;  
RA Footz T.K., Brinkman-Wills P., Banting G.S., Maier S.A., Riaz M.A.,  
RA Bridgland L.J., Hu S., Barren B., Minoshima S., Shimizu N., Pan H.,  
RA Nguyen T., Fang F., Fu Y., Ray L., Wu H., Shaull S., Phan S., Yao Z.,  
RA Chen F., Huan A., Hu P., Wang Q., Loh P., Qi S., Roe B.A.,  
RA McDermid H.E.;  
RT "Analysis of the cat eye syndrome critical region in humans and the  
RT region of conserved synteny in mice: a search for candidate genes at  
RT or near the human chromosome 22 pericentromere.";  
RL Genome Res. 11:1053-1070(2001).  
CC -!- TISSUE SPECIFICITY: Adult heart and skeletal muscle. Widely  
CC expressed in fetal tissues.  
CC -!- MISCELLANEOUS: Candidate gene for the Cat Eye Syndrome (CES), a  
CC developmental disorder associated with the duplication of a 2 Mb  
CC region of 22q11.2. Duplication usually takes in the form of a  
CC supernumerary bisatellited isodicentric chromosome, resulting in  
CC four copies of the region (represents an inv dup(22)(q11)). CES is  
CC characterized clinically by the combination of coloboma of the  
CC iris and anal atresia with fistula, downslanting palpebral  
CC fissures, preauricular tags and/or pits, frequent occurrence of  
CC heart and renal malformations, and normal or near-normal mental  
CC development.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF307448; AAK30048.1; -;  
CC Genbank; HGNC:1842; CECR4.  
CC NON TER 1  
SQ SEQUENCE 124 AA; 12838 MW; 938E00386308ECTA CRC64;  
Query Match 100.0%; Score 28; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 RGRPQ 5  
Db 58 RGRPQ 62  
RESULT 3  
ID PA23 TRIOK STANDARD; PRT; 137 AA.  
AC Q92152;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE Phospholipase A2 homolog PLA2-03 precursor.  
OS Trimeresurus okinavensis (Hime-habu) (Ovophis okinavensis).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
OC Viperidae; Crotalinae; Ovophis.  
OX NCBI\_TaxID=8769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Venom gland;  
RX MEDLINE=96269416; PubMed=8682315;  
RA Nobuhisa I., Nakashima K.-I., Deshimaru M., Ogawa T., Shimohigashi Y.,  
RA Fukumaki Y., Sakaki Y., Hattori S., Kihara H., Ohno M.;  
RT "Accelerated evolution of Trimeresurus okinavensis venom gland  
RT phospholipase A2 isozyme-encoding genes.";  
RL Gene 172:267-272(1996).

CC -!- FUNCTION: Lacks PA2 enzymatic activity (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
CC -!- MISCELLANEOUS: Does not bind calcium as one of the calcium binding  
CC ligands is lost (Asp->Lys in position 64) (By similarity).  
CC -!- SIMILARITY: Belongs to the phospholipase A2 family. Group II  
CC subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D49389; BAA08384.1; -;  
CC FRL; JC4877; JC4877.  
CC HSP; P81165; IG0D.  
CC InterPro; IPR001211; PhospholipaseA2.  
CC Pfam; PF00068; phoslip; 1.  
CC PRINTS; PR00389; PPHLIPASEA2.  
CC ProDom; PD000303; PhospholipaseA2; 1.  
CC SMART; SM00085; PA2C; 1.  
CC PROSITE; PS00119; PA2\_ASP; 1.  
CC PROSITE; PS00118; PA2\_HIS; 1.  
KW SIGNAL.  
FT SIGNAL 1 16 BY SIMILARITY.  
FT CHAIN 17 137 PHOSPHOLIPASE A2 HOMOLOG PLA2-03.  
FT ACT\_SITE 63 63 BY SIMILARITY.  
FT ACT\_SITE 105 105 BY SIMILARITY.  
FT DISULFID 42 131 BY SIMILARITY.  
FT DISULFID 44 60 BY SIMILARITY.  
FT DISULFID 59 111 BY SIMILARITY.  
FT DISULFID 65 137 BY SIMILARITY.  
FT DISULFID 66 104 BY SIMILARITY.  
FT DISULFID 73 97 BY SIMILARITY.  
FT DISULFID 91 102 BY SIMILARITY.  
SQ SEQUENCE 137 AA; 15453 MW; FC8D15BB764E45E CRC64;  
Query Match 100.0%; Score 28; DB 1; Length 137;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 RGRPQ 5  
Db 49 RGRPQ 53  
RESULT 4  
ID NEU3 ONCKE STANDARD; PRT; 153 AA.  
AC P16041; Q91173;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Vasotocin-neurophysin VT 1 precursor [Contains: Vasotocin (VT);  
DE Neurophysin VT 1].  
OS Eukaryotus Keta (Chum salmon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
OX NCBI\_TaxID=8018;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=90127450; PubMed=2298304;  
RA Heierhorst J., Mahlmann S., Morley S.D., Coe I.R., Sherwood N.M.,  
RA Richter D.;  
RT "Molecular cloning of two distinct vasotocin precursor cDNAs from  
RT chum salmon (Oncorhynchus keta) suggests an ancient gene  
RT duplication.";  
RL FEBS Lett. 260:301-304(1990).

```

RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Taugaruihi, TISSUE=Brain;
RX  MEDLINE=91258631; PubMed=2045542;
RA  Hyodo S., Kato Y., Ono M., Urano A.;
RT  "Cloning and sequence analyses of cDNAs encoding vasotocin and
RT  isotocin precursors of chum salmon, Oncorhynchus keta: evolutionary
RL  J. Comp. Physiol. B 140:601-608(1991).
CC  -!- FUNCTION: Vasotocin is an antidiuretic hormone.
CC  -!- SUBCELLULAR LOCATION: Secreted.
CC  -!- PM: In non-mammalian tetrapods, the proteolytic processing of the
CC  pro-vasotocin involves only one cleavage, releasing the hormone
CC  moiety and a "big" neurophysin with two domains homologous to the
CC  mammalian neurophysin II and copeptin, respectively.
CC  -!- PM: Seven disulfide bonds are present in neurophysin.
CC  -!- SIMILARITY: Belongs to the vasopressin/oxytocin family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  at the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X17327; CAA35205.1; -
DR  EMBL; D10942; BAA01736.1; -
DR  PIR; A34132; A34132.
DR  HSSP; P01180; INPO.
DR  InterPro; IPR000981; Neurohyp_horm.
DR  Pfam; PF00220; hormones; 1.
DR  Pfam; PF00184; hormones; 1.
DR  PRINTS; PR00831; NEUROPHYSIN.
DR  ProDom; PD001676; Neurohyp_horm; 1.
DR  SMART; SM00003; NH; 1.
DR  PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW  Hormone, Hypothalamus, Amidation; Cleavage on pair of basic residues;
KW  Multigene family; Signal.
FT  SIGNAL 1 20
FT  PEPTIDE 21 29 VASOTOCIN.
FT  PEPTIDE 33 153 NEUROPHYSIN VT 1.
FT  DISULFID 21 26 BY SIMILARITY.
FT  MOD RES 29 29 AMIDATION (G-30 PROVIDE AMIDE GROUP).
FT  CONFLICT 7 7 P -> Q (IN REF. 2).
SQ  SEQUENCE 153 AA; 15995 MW; A67C051358E4D5A9 CRC64;
Query Match 100.0%; Score 28; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRPQ 5
Db 149 RGRPQ 153
|||||
RESULT 5
PRPC_HUMAN
ID PRPC_HUMAN STANDARD; PRT; 166 AA.
AC P02810;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-
DE 3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains:
DE Peptide P-C].
GN PRH1 AND PRH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).

```

```

RX  MEDLINE=86196106; PubMed=3009472;
RA  Kim H.-S., Maeda N.;
RT  "Structures of two HaeIII-type genes in the human salivary
RT  proline-rich protein multigene family.";
RL  J. Biol. Chem. 261:6712-6718(1986).
RN  [2]
RP  SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).
RX  MEDLINE=85289325; PubMed=2993301;
RA  Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT  "Differential RNA splicing and post-translational cleavages in the
RT  human salivary proline-rich protein gene system.";
RL  J. Biol. Chem. 260:11123-11130(1985).
RN  [3]
RP  SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).
RX  MEDLINE=89061650; PubMed=3196309;
RA  Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K.,
RA  Madapallimattam G., Schluckebier S.K.;
RT  "The primary structures of six human salivary acidic proline-rich
RT  proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";
RL  Biochem. J. 255:15-21(1988).
RN  [4]
RP  SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).
RX  MEDLINE=88074309; PubMed=3687941;
RA  Azen E.A., Kim H.-S., Goodman P., Flynn S., Maeda N.;
RT  "Alleles at the PRH1 locus coding for the human salivary-acidic
RT  proline-rich proteins Pa, Db, and PIF.";
RL  Am. J. Hum. Genet. 41:1035-1047(1987).
RN  [5]
RP  SEQUENCE OF 17-166 (PRP-2).
RX  MEDLINE=86222916; PubMed=3710693;
RA  Schlesinger D.H., Hay D.I.;
RT  "Complete covalent structure of a proline-rich phosphoprotein, PRP-2,
RT  an inhibitor of calcium phosphate crystal growth from human parotid
RT  saliva.";
RL  Int. J. Pept. Protein Res. 27:373-379(1986).
RN  [6]
RP  SEQUENCE OF 17-166 (PROTEIN C).
RX  MEDLINE=80204368; PubMed=7300845;
RA  Wong R.S.C., Bennick A.;
RT  "The primary structure of a salivary calcium-binding proline-rich
RT  phosphoprotein (protein C), a possible precursor of a related
RT  salivary protein A.";
RL  J. Biol. Chem. 255:5943-5948(1980).
RN  [7]
RP  SEQUENCE OF 17-46 (PROTEIN C).
RX  MEDLINE=81191179; PubMed=7228490;
RA  Schlesinger D.H., Hay D.I.;
RT  "Primary structure of the active tryptic fragments of human and
RT  monkey salivary anionic proline-rich proteins.";
RL  Int. J. Pept. Protein Res. 17:34-41(1981).
RN  [8]
RP  SEQUENCE OF 17-122 (PROTEIN A).
RX  MEDLINE=79173237; PubMed=438215;
RA  Wong R.S.C., Hofmann T., Bennick A.;
RT  "The complete primary structure of a proline-rich phosphoprotein from
RT  human saliva.";
RL  J. Biol. Chem. 254:4800-4808(1979).
RN  [9]
RP  SEQUENCE OF 17-122 (PROTEIN A).
RA  Schlesinger D.H., Hay D.I.;
RT  "Complete primary structure of a proline-rich phosphoprotein (PRP-4),
RT  a potent inhibitor of calcium phosphate precipitation in human parotid
RT  saliva.";
RL  (In) Gross E., Meienhofer J. (eds.);
RL  Peptides: structure and biological function (Proceedings of the 6th
RL  American peptide symposium), pp.133-136, Pierce Chemical Co.,
RL  Rockford IL. (1979).
RN  [10]
RP  SEQUENCE OF 123-166 (PEPTIDE P-C).
EX  MEDLINE=80227634; PubMed=7390979;
RA  Isemura S., Saitoh E., Sanada K.;
RT  "The amino acid sequence of a salivary proline-rich peptide, P-C, and
RT  its relation to a salivary proline-rich phosphoprotein, protein C.";

```

```

RL J. Biochem. 87:1071-1077(1980).
RN [1]
RP VARIANT PRH2-3 LYS-163.
RA Azen E.A.;
RT "A frequent mutation in the acidic proline-rich protein gene, PRH2,
RT causing a Q47K change closely adjacent to the bacterial binding
RT domain of the cognate salivary PRP (Pr1)' in Afro-Americans.";
RL Hum. Mutat. 12:72-72(1998).
CC -1- FUNCTION: PRP's act as highly potent inhibitors of crystal growth
CC of calcium phosphates. They provide a protective and reparative
CC environment for dental enamel which is important for the integrity
CC of the teeth.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIF-S yield PRP-4,
CC PRP-3 (protein A), and PIF-F, respectively.
CC -1- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele;
CC allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is
CC also known as PR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03202; AAA60183.1; -
DR EMBL; K03203; AAA60184.1; -
DR EMBL; M13057; AAA98807.1; -
DR EMBL; M13058; AAA98808.1; -
DR Genew; HGNC:9366; PRH1.
DR Genew; HGNC:9367; PRH2.
DR MIM; 168730; -
DR MIM; 168790; -
DR MIM; 168710; -
DR GO; GO:0005615; C:extracellular space; TAS.
KW Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism;
KW Pyroglutamate carboxylic acid.
FT SIGNAL 1 16
FT CHAIN 17 166
FT FT CHAIN 17 122
FT FT CHAIN 123 166
FT FT CHAIN 17 46
FT FT CHAIN 17 17
FT FT MOD RES 24 24
FT FT MOD RES 38 38
FT FT VARIANT 20 20
FT FT VARIANT 66 66
FT FT VARIANT 163 163
FT FT CONFLICT 41 41
FT FT SEQUENCE 166 AA; 17017 MW; A7DF62BF94E33CEF CRC64;
SQ
Query Match 100.0%; Score 28; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 RGRPQ 5
DB 122 RGRPQ 126
RESULT 6
ID YP33 METTF STANDARD; PRT; 197 AA.
AC P29573;
DT 01-APR-1993 (Rel. 25, Created)

```

---

```

DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Hypothetical 22.5 kDa protein (ORF3).
OS Methanobacterium thermoformicicum.
OG Plasmid pFVI.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3848 / THE;
RX MEDLINE=93126090; PubMed=1336177;
RA Noelling J., van Breden F.J.M., Eggen R.I.L., de Vos W.M.;
RT "Modular organization of related Archaeal plasmids encoding different
RT restriction-modification systems in Methanobacterium
RT thermoformicicum.";
RL Nucleic Acids Res. 20:6501-6507(1992).
CC -1- SIMILARITY: TO THE CORRESPONDING ORF IN PF21.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68366; CAA48427.1; -
DR PIR; S30303; S26438.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 197 AA; 22467 MW; 5B8A0D9444E106A3 CRC64;
Query Match 100.0%; Score 28; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 RGRPQ 5
DB 118 RGRPQ 122
RESULT 7
ID YP23 METTF STANDARD; PRT; 217 AA.
AC P29574;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Hypothetical 24.9 kDa protein (ORF3').
OS Methanobacterium thermoformicicum.
OG Plasmid pF21.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3720 / Z-245;
RX MEDLINE=93126090; PubMed=1336177;
RA Noelling J., van Breden F.J.M., Eggen R.I.L., de Vos W.M.;
RT "Modular organization of related Archaeal plasmids encoding different
RT restriction-modification systems in Methanobacterium
RT thermoformicicum.";
RL Nucleic Acids Res. 20:6501-6507(1992).
CC -1- SIMILARITY: TO THE CORRESPONDING ORF IN PFV1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```



```

DR EMBL; X68367; CAA48442.1; -.
DR PIR; S30318; S26452.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 217 AA; 24950 MW; 7FA6AA3B3A989569 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5
DB 138 RGRPQ 142

RESULT 8
TRA2_ECOLI STANDARD; PRT; 307 AA.
AC P03009;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Probable transposase for transposon Tn903.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN Nucleotide sequence of the kanamycin resistance transposon Tn903.;
RP SEQUENCE FROM N.A.
RX MEDLINE=82033200; PubMed=6270337;
RA Oka A., Sugisaki H., Takanami M.;
RT "Nucleotide sequence of the kanamycin resistance transposon Tn903.";
RL J. Mol. Biol. 147:217-226(1981).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=81175113; PubMed=6261245;
RA Grindley N.D.F., Joyce C.M.;
RT "Genetic and DNA sequence analysis of the kanamycin resistance transposon Tn903.";
RL Proc. Natl. Acad. Sci. U.S.A. 77:7176-7180(1980).
RN (3)
RP SEQUENCE FROM N.A.
RX MEDLINE=82049485; PubMed=6271455;
RA Grindley N.D.F., Joyce C.M.;
RT "Analysis of the structure and function of the kanamycin-resistance transposon Tn903.";
RL Cold Spring Harb. Symp. Quant. Biol. 45:125-133(1981).
CC -!- FUNCTION: REQUIRED FOR TRANSDUCTION OF TRANSPOSON Tn903.
CC -!- MISCELLANEOUS: TRANSPOSON Tn903 CODES FOR KANAMYCIN RESISTANCE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; V00621; CAA23897.1; -.
DR EMBL; V00359; CAA23657.1; -.
DR EMBL; M10597; AAA27447.1; -.
DR PIR; A92864; TQEC93.
DR InterPro; IPR002559; Transposase_11.
DR Pfam; PF01609; Transposase_11.
KW Transposable element; Transposition; DNA-binding; DNA recombination.
SQ SEQUENCE 307 AA; 34933 MW; 8154829104E06980 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 307;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5
DB 45 RGRPQ 49

RESULT 9
RLUD_RALSO STANDARD; PRT; 358 AA.
AC Q8YX8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
DE (Pseudouridylate synthase) (Uracil hydrolyase).
GN RLUD OR RSC1628 OR RS03991.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GMI1000;
RC MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Arlat M., Billault A., Brottier P., Camus J.C., Catolico L., Chandelier M., Choisme N., Claudel-Renard C., Cunnac S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T., Sigulier P., Thebault P., Whalen M., Wincker P., Levy M., Weissenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RT Nature 415:497-502(2002).
RL -!- FUNCTION: Responsible for synthesis of pseudouridine from uracil at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By similarity).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine 5'-phosphate + H(2)O.
CC -!- SIMILARITY: Belongs to the pseudouridine synthase rluA family.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; AL646065; CAD15330.1; -.
DR InterPro; IPR006225; Pseud_synth_RluD.
DR InterPro; IPR006145; Pseudou synth.
DR InterPro; IPR006224; Rlu_synth.
DR InterPro; IPR002942; S4.
DR Pfam; PF00849; Pseudou_synth_2; 1.
DR Pfam; PF01479; S4; 1.
DR ProDom; PD001819; PSI_RLU; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00005; rluD subfam; 1.
DR PROSITE; PS01129; PSI_RLU; 1.
DR PROSITE; PS50839; S4; 1.
KW RNA processing; Lyase; RNA-binding; Complete proteome.
FT DOMAIN 50 114 S4 RNA-BINDING.
FT ACT_SITE 171 171 BY SIMILARITY.
SQ SEQUENCE 358 AA; 38687 MW; 0A11CDEA474BFB5A CRC64;

Query Match 100.0%; Score 28; DB 1; Length 358;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5
DB 297 RGRPQ 301

RESULT 10
FADH_ARYME STANDARD; PRT; 360 AA.
ID FADH_ARYME

```

P80094;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE NAD/mycothiol-dependent formaldehyde dehydrogenase (EC 1.2.1.66) (MD-  
 DE FALDH1)  
 OS Amycolatopsis methanolica.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.  
 OX NCBI\_TaxID=1814;  
 RN [1]  
 RP SEQUENCE  
 RX MEDLINE=98004265; PubMed=9346279;  
 RA Jorin A.; van Ophem P.W.; Piersma S.R.; Persson B.; Duine J.A.;  
 RA Joernvall H.;  
 RT "Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic  
 RT medium-chain dehydrogenase/reductase, phylogenetically links  
 RT different eukaryotic alcohol dehydrogenases-- primary structure,  
 RT conformational modelling and functional correlations.";  
 RL Eur. J. Biochem. 248:282-289 (1997).  
 RN [2]  
 RP PRELIMINARY SEQUENCE OF 1-31.  
 RX MEDLINE=92282378; PubMed=1597190;  
 RA van Ophem P.W.; van Beumen J.; Duine J.A.;  
 RT "NAD-linked, factor-dependent formaldehyde dehydrogenase or trimeric,  
 RT zinc-containing, long-chain alcohol dehydrogenase from Amycolatopsis  
 RT methanolica.";  
 RL Eur. J. Biochem. 206:511-518 (1992).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=97345657; PubMed=9202149;  
 RA Misset-Smiths M.; van Ophem P.W.; Sakuda S.; Duine J.A.;  
 RT "Mycothiol, 1-O-(2'-[N-acetyl-L-cysteine]methylamido-2'-deoxy-alpha-  
 RT D-glucopyranosyl)-D-myo-inositol, is the factor of NAD/factor-  
 RT dependent formaldehyde dehydrogenase.";  
 RL FEBS Lett. 409:221-222 (1997).  
 CC -!- CATALYTIC ACTIVITY: Formaldehyde + mycothiol + NAD(+) = S-  
 CC forylmycothiol + NADH.  
 CC -!- COPACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -!- SUBUNIT: Homotrimer.  
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.  
 DR InterPro: IPR002328; ADH\_zinc.  
 DR InterPro: IPR002085; Adh\_zn family.  
 DR InterPro: IPR00205; NAD\_BS.  
 DR Pfam: PF00107; ADH\_zinc\_N; 1.  
 DR PROSITE: PS00059; ADH\_ZINC; 1.  
 KW Oxidoreductase; Zinc; Metal-binding; NAD.  
 FT ACT SITE 89 89 SUBSTRATE BINDING (PROBABLE).  
 FT METAL 41 41 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 62 62 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 92 92 ZINC 2 (BY SIMILARITY).  
 FT METAL 95 95 ZINC 2 (BY SIMILARITY).  
 FT METAL 98 98 ZINC 2 (BY SIMILARITY).  
 FT METAL 106 106 ZINC 2 (BY SIMILARITY).  
 FT METAL 161 161 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 360 AA; 37757 MW; 48BD1F7CEAB2CB7 CRC64;  
 Query Match 100.0%; Score 28; DB 1; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPQ 5  
 Db 100 RGRPQ 104  
 RESULT 11  
 AZAB ELENA  
 ID AZAB ELENA STANDARD; PRT; 384 AA.  
 AC O19014;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor) (Fragment).  
 GN ADRA2B.  
 OS Elephas maximus (Indian elephant).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.  
 OX NCBI\_TaxID=9783;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97357151; PubMed=9214502;  
 RA Springer M.S.; Cleven G.C.; Madsen O.J.; de Jong W.W.; Waddell V.G.;  
 RA Amrine H.M.; Stanhope M.J.;  
 RT "Endemic African mammals shake the phylogenetic tree.";  
 RL Nature 388:61-64 (1997).  
 CC -!- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-  
 CC induced inhibition of adenylate cyclase through the action of G  
 CC proteins.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Y12525; CAA73125.1; -.  
 DR HSPP; P29274; LMMH.  
 DR InterPro: IPR000276; GPCR\_Rhodopsn.  
 DR Pfam: PF00001; 7tm\_1; 1.  
 DR PRINTS; PR00237; GPCRHOPOPSN.  
 DR PROSITE; PS00237; G PROTEIN RECEPTOR FL1; 1.  
 DR PROSITE; PS00262; G PROTEIN RECEPTOR FL2; 1.  
 KW G-protein coupled receptor; Transmembrane; Multigene family;  
 KW Phosphorylation; Lipoprotein; Palmitate.  
 FT NON TER 1 1  
 FT TRANSMEM <1 25 1 (POTENTIAL).  
 FT DOMAIN 26 36 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 37 62 2 (POTENTIAL).  
 FT DOMAIN 63 72 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 73 95 3 (POTENTIAL).  
 FT DOMAIN 96 117 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 118 140 4 (POTENTIAL).  
 FT DOMAIN 141 156 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 157 180 5 (POTENTIAL).  
 FT DOMAIN 181 348 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 349 372 6 (POTENTIAL).  
 FT DOMAIN 373 381 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 382 >384 7 (POTENTIAL).  
 FT DISULFID 72 151 BY SIMILARITY.  
 FT DOMAIN 281 285 ASP/GLU-RICH (ACIDIC).  
 FT SITE 79 79 IMPLICATED IN LIGAND BINDING (BY  
 FT SIMILARITY).  
 FT SITE 163 163 IMPLICATED IN CATECHOL AGONIST BINDING  
 FT (BY SIMILARITY).  
 FT SITE 167 167 IMPLICATED IN CATECHOL AGONIST BINDING  
 FT (BY SIMILARITY).  
 FT NON TER 384 384  
 SQ SEQUENCE 384 AA; 41911 MW; CF41B56CC355B94F CRC64;  
 Query Match 100.0%; Score 28; DB 1; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPQ 5  
 Db 146 RGRPQ 150  
 RESULT 12  
 AZAB\_MACPR

ID A2AB MACPR STANDARD; PRT; 387 AA.  
 AC O19025;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor) (Fragment).  
 GN ADRA2B.  
 OS Macroscelides proboscideus (Short-eared elephant shrew).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Macroscelidea; Macroscelididae; Macroscelididae.  
 OX NCBI\_TaxID=29082;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=97357151; PubMed=9214502;  
 RA Springer M.S., Cleven G.C., Madsen O.J., de Jong W.W., Waddell V.G.,  
 RA Arine H.M., Stanhope M.J.;  
 RT "Endemic African mammals shake the phylogenetic tree.";  
 RL Nature 388:61-64(1997).  
 RN [2]  
 RP REVISIONS TO 148 AND 255.  
 RA Springer M.S., Cleven G.C., Madsen O.J., de Jong W.W., Waddell V.G.,  
 RA Arine H.M., Stanhope M.J.;  
 RA Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 RL CC -1- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-  
 CC induced inhibition of adenylate cyclase through the action of G  
 CC proteins.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y12524; CAAT73124.2; -;  
 DR HSP; P29274; 1MMH.  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm.1; 1.  
 DR PRINTS; PR00237; GPCRHOPOPSN.  
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.  
 DR PROSITE; PS00262; G-PROTEIN RECEPTOR FL2; 1.  
 KW G-protein coupled receptor; Transmembrane; Multigene family;  
 KW Phosphorylation; Lipoprotein; Palmitate.  
 FT NON\_TER 1 1  
 FT TRANSMEM <1 25  
 FT DOMAIN 26 36  
 FT DOMAIN 37 62  
 FT TRANSMEM 63 72  
 FT TRANSMEM 73 95  
 FT TRANSMEM 96 117  
 FT TRANSMEM 118 140  
 FT TRANSMEM 141 156  
 FT TRANSMEM 157 180  
 FT TRANSMEM 181 351  
 FT TRANSMEM 352 375  
 FT TRANSMEM 376 384  
 FT TRANSMEM 385 >387  
 FT TRANSMEM 280 288  
 FT TRANSMEM 72 151  
 FT TRANSMEM 79 79  
 FT SITE 163 163  
 FT SITE 167 167  
 FT SITE 387 387  
 FT NON\_TER 387 387  
 FT SEQUENCE 387 AA; 42587 MW; B74AD5F0EE23BD5A CRC64;  
 SQ  
 Query Match 100.0%; Score 28; DB 1; Length 387;  
 Best Local Similarity 100.0%; Pred. No. 52;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RORPQ 5  
 DB 146 RGRPQ 150  
 RESULT 13  
 A2AB ORYAF STANDARD; PRT; 388 AA.  
 ID A2AB ORYAF  
 AC O19032;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor) (Fragment).  
 GN ADRA2B.  
 OS Oryzomys afer (Aardvark).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Tubulidentata; Orycteropodidae; Orycteropus.  
 OX NCBI\_TaxID=9818;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=97357151; PubMed=9214502;  
 RA Springer M.S., Cleven G.C., Madsen O.J., de Jong W.W., Waddell V.G.,  
 RA Arine H.M., Stanhope M.J.;  
 RT "Endemic African mammals shake the phylogenetic tree.";  
 RL Nature 388:61-64(1997).  
 RN CC -1- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-  
 CC induced inhibition of adenylate cyclase through the action of G  
 CC proteins.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y12522; CAAT73122.2; ALT SEQ.  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm.1; 1.  
 DR PRINTS; PR00237; GPCRHOPOPSN.  
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.  
 DR PROSITE; PS00262; G-PROTEIN RECEPTOR FL2; 1.  
 KW G-protein coupled receptor; Transmembrane; Multigene family;  
 KW Phosphorylation; Lipoprotein; Palmitate.  
 FT NON\_TER 1 1  
 FT TRANSMEM <1 25  
 FT DOMAIN 26 36  
 FT TRANSMEM 37 62  
 FT TRANSMEM 63 72  
 FT TRANSMEM 73 95  
 FT TRANSMEM 96 117  
 FT TRANSMEM 118 140  
 FT TRANSMEM 141 156  
 FT TRANSMEM 157 180  
 FT TRANSMEM 181 352  
 FT TRANSMEM 353 376  
 FT TRANSMEM 377 385  
 FT TRANSMEM 386 >388  
 FT TRANSMEM 72 151  
 FT TRANSMEM 79 79  
 FT SITE 163 163  
 FT SITE 167 167  
 FT SITE 388 388  
 FT NON\_TER 388 388  
 FT SEQUENCE 388 AA; 42429 MW; 1810DC767E838897 CRC64;  
 SQ

Query Match 100.0%; Score 28; DB 1; Length 388;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
 DB 146 RGRPQ 150

RESULT 14  
 A2AB\_HORSE STANDARD; PRT; 389 AA.  
 ID A2AB\_HORSE  
 AC O77721;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor) (Fragment).  
 GN ADRA2B.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98334573; PubMed=9667998;  
 RA Stanhope M.J., Madsen O.J., Waddell V.G., de Jong W.W.,  
 RA Springer M.S.;  
 RT "Highly congruent molecular support for a diverse superordinal clade  
 of endemic African mammals.";  
 RL Mol. Phylogenet. Evol. 9:501-508(1998).  
 RN [2]  
 RN REVISIONS.  
 RA Madsen O.J.;  
 RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RL CC -1- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-  
 induced inhibition of adenylate cyclase through the action of G  
 proteins.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Y15945; CAA75898.2; --  
 CC HSP; P29274; LMWH.  
 CC InterPro; IPR000276; GPCR\_Rhodpsn.  
 CC Pfam; PF00001; 7tm.1; 1.  
 CC PRINTS; PS00237; GPCR\_Rhodopsin.  
 CC PROSITE; PS00237; G-PROTEIN\_RECEP\_F1\_1; 1.  
 CC PROSITE; PS00262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 CC G-protein coupled receptor; Transmembrane; Multigene family.  
 CC NON\_TER 1  
 CC TRANSMEM <1 25  
 CC DOMAIN 26 36  
 CC TRANSMEM 37 62  
 CC DOMAIN 63 72  
 CC TRANSMEM 73 95  
 CC DOMAIN 96 117  
 CC TRANSMEM 118 140  
 CC DOMAIN 141 156  
 CC TRANSMEM 157 180  
 CC DOMAIN 181 363  
 CC TRANSMEM 364 387  
 CC TRANSMEM 388 >389  
 CC DOMAIN 389  
 CC TRANSMEM 72 151  
 CC DISULFID 281 302  
 CC DOMAIN 281 302  
 CC SITE 79 IMPLICATED IN LIGAND BINDING (BY

FT SITE 163 163  
 FT SITE 167 167  
 FT NON\_TER 389 389  
 SQ SEQUENCE 389 AA; 42257 MW; 992179431679B0FD CRC64;  
 Query Match 100.0%; Score 28; DB 1; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
 DB 146 RGRPQ 150

RESULT 15  
 A2AB\_PROHA STANDARD; PRT; 389 AA.  
 ID A2AB\_PROHA  
 AC O19054;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor) (Fragment).  
 GN ADRA2B.  
 OS Procavia capensis habessinica (Abyssinian hyrax).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Hyracoidea; Procaviidae; Procavia.  
 OX NCBI\_TaxID=9814;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=97357151; PubMed=9214502;  
 RA Springer M.S., Cleven G.C., Madsen O.J., de Jong W.W., Waddell V.G.,  
 RA Amrine H.M., Stanhope M.J.;  
 RT "Endemic African mammals shake the phylogenetic tree.";  
 RL Nature 388:61-64(1997).  
 CC -1- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-  
 induced inhibition of adenylate cyclase through the action of G  
 proteins.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Y12523; CAA73123.1; --  
 CC HSP; P29274; LMWH.  
 CC InterPro; IPR000276; GPCR\_Rhodpsn.  
 CC Pfam; PF00001; 7tm.1; 1.  
 CC PRINTS; PS00237; GPCR\_Rhodopsin.  
 CC PROSITE; PS00237; G-PROTEIN\_RECEP\_F1\_1; 1.  
 CC PROSITE; PS00262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 CC G-protein coupled receptor; Transmembrane; Multigene family;  
 CC Phosphorylation; Lipoprotein; Palmitate.  
 CC NON\_TER 1  
 CC TRANSMEM <1 25  
 CC DOMAIN 26 36  
 CC TRANSMEM 37 62  
 CC DOMAIN 63 72  
 CC TRANSMEM 73 95  
 CC DOMAIN 96 117  
 CC TRANSMEM 118 140  
 CC DOMAIN 141 156  
 CC TRANSMEM 157 180  
 CC DOMAIN 181 353  
 CC TRANSMEM 354 377  
 CC DOMAIN 378 396  
 CC SITE 396

PT TRANSMEM 387 >389 7 (POTENTIAL).  
 PT DOMAIN 281 292 ASP/GLU-RICH (ACIDIC).  
 FT DISULFID 72 151 BY SIMILARITY.  
 FT SITE 79 79 IMPLICATED IN LIGAND BINDING (BY  
 SIMILARITY).  
 FT SITE 163 163 IMPLICATED IN CATECHOL AGONIST BINDING  
 (BY SIMILARITY).  
 FT SITE 167 167 IMPLICATED IN CATECHOL AGONIST BINDING  
 (BY SIMILARITY).  
 FT NON\_TER 389 389  
 SQ SEQUENCE 389 AA; 42528 MW; 4F2089EA140976E1 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RGRPQ 5  
 Db 146 RGRPQ 150

Search completed: April 6, 2004, 16:07:57  
 Job time : 4.2243 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 18.0841 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-1

Perfect score: 28

Sequence: 1 RGRPQ 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	54	2 Q9X4U1	Q9x4u1 klebsiella
2	28	100.0	54	10 Q7X859	Q7x859 oryza sativ
3	28	100.0	84	5 Q9V856	Q9v856 drosophila
4	28	100.0	87	16 Q89NY2	Q89ny2 bradyrhizob
5	28	100.0	91	2 Q52183	Q52183 synechocyst
6	28	100.0	122	4 Q812S3	Q812s3 homo sapien
7	28	100.0	130	10 Q7XBG3	Q7xbg3 setaria ita
8	28	100.0	134	10 Q8S1G6	Q8s1g6 oryza sativ
9	28	100.0	138	10 Q239V7	Q239v7 helianthus
10	28	100.0	138	10 Q8RVV7	Q8rvv7 helianthus
11	28	100.0	140	4 Q8NDY4	Q8ndy4 homo sapien
12	28	100.0	141	10 Q48526	Q48526 arabidopsis
13	28	100.0	141	10 Q9MZQ5	Q9mzq5 arabidopsis
14	28	100.0	142	5 Q86NA0	Q86na0 drosophila
15	28	100.0	142	10 Q8LFA8	Q8lfa8 arabidopsis
16	28	100.0	149	16 Q889R1	Q889r1 pseudomonas

17	28	100.0	157	16 Q89YE0	Q89ye0 bradyrhizob
18	28	100.0	164	10 Q943B9	Q943b9 oryza sativ
19	28	100.0	165	5 Q22835	Q22835 caenorhabdi
20	28	100.0	177	16 Q9RRY8	Q9rry8 deinococcus
21	28	100.0	179	17 Q9YAJ0	Q9yaj0 aeropyrum p
22	28	100.0	182	16 Q98M89	Q98m89 rhizobium l
23	28	100.0	194	12 Q919T5	Q919t5 influenza a
24	28	100.0	209	4 Q8N811	Q8n811 homo sapien
25	28	100.0	213	11 Q8CAS4	Q8cas4 mus musculu
26	28	100.0	215	2 Q9KIH5	Q9kih5 rhizobium e
27	28	100.0	215	10 Q9AWL3	Q9awl3 oryza sativ
28	28	100.0	221	10 Q8W0X7	Q8w0x7 oryza sativ
29	28	100.0	225	16 Q9RVN8	Q9rvn8 deinococcus
30	28	100.0	227	16 Q89XI3	Q89xi3 bradyrhizob
31	28	100.0	234	7 Q8XBE4	Q8xre4 brachydanio
32	28	100.0	237	13 Q9PUT2	Q9put2 brachydanio
33	28	100.0	249	11 Q922D9	Q922d9 mus musculu
34	28	100.0	254	12 Q91TN4	Q91tn4 tupala herp
35	28	100.0	256	16 Q7UYG7	Q7uyg7 rhodopirell
36	28	100.0	262	10 Q9FSV4	Q9fsv4 fagus sylv
37	28	100.0	267	16 Q825X6	Q825x6 streptomyc
38	28	100.0	276	11 Q80XY0	Q80xy0 rattus sp.
39	28	100.0	278	2 Q860S9	Q860s9 acospirillu
40	28	100.0	280	2 Q7WZ24	Q7wz24 pseudomonas
41	28	100.0	289	17 Q9YBY9	Q9yby9 aeropyrum p
42	28	100.0	301	16 Q9RX03	Q9rx03 deinococcus
43	28	100.0	301	16 Q9RWL5	Q9rwl5 deinococcus
44	28	100.0	307	2 Q48345	Q48345 escherichia
45	28	100.0	307	2 Q939K1	Q939k1 klebsiella

#### ALIGNMENTS

##### RESULT 1

Q9X4U1 PRELIMINARY; PRT; 54 AA.  
ID Q9X4U1  
AC Q9X4U1;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE IS903 homolog (Fragment).  
OS Klebsiella pneumoniae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Klebsiella.  
OX NCBI\_TaxID=573;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=89/50;  
RX MEDLINE=99194747; PubMed=10094716;  
RA Rahn A., Drummelsmith J., Whitfield C.;  
RT "Conserved organization in the cps gene clusters for expression of  
RT Escherichia coli group 1 K antigens: relationship to the colanic acid  
RT biosynthesis locus and the cps genes from Klebsiella pneumoniae.";  
RL J. Bacteriol. 181:2307-2313(1999).  
DR EMBL; AF18259; AAD30025.1; -.  
FT NON-TER  
SQ SEQUENCE 54 AA; 6322 MW; 2CD59CAB56837D06 CRC64;

Query Match 100.0%; Score 28; DB 2; Length 54;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
|||  
Db 45 RGRPQ 49

##### RESULT 2

Q7X859 PRELIMINARY; PRT; 54 AA.  
ID Q7X859  
AC Q7X859;  
DT 01-OCT-2003 (TREMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein P001H09.116 (Hypothetical protein  
 DE P0506F02.130)  
 GN P001H09.116 OR P0506F02.130  
 OS Oryza sativa (japonica cultivar-group)  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoideae; Oryza.  
 ON NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RC Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GAS) genomic DNA, chromosome 7, PAC  
 RT clone:P001H09.1;"  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RC Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GAS) genomic DNA, chromosome 7, PAC  
 RT clone:P0506F02.1;"  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AP004260; BAC79724.1; -  
 DR EMBL; AP004306; BAC79861.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 54 AA; 5668 MW; 4BDDC60331A4F60A CRC64;  
  
 Query Match 100.0%; Score 28; DB 10; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRPQ 5  
 DB 41 RGRPQ 45  
  
 RESULT 3  
 Q9V856  
 ID Q9V856 PRELIMINARY; PRT; 84 AA.  
 AC Q9V856  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CG18186 protein.  
 GN CG18186  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Venter L.M., Zhang O., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.A., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Chadiet E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hoskin D., Houston K.A., Howland T.O., Wei M.-H., Ibegwam C.,  
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei E., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulyov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M., P.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RL EMBL; AE003802; AAF57821.1; -  
 DR HSSP; P23197; IAP0.  
 DR FlyBase; FBgn0034255; CG18186.  
 DR GO; GO:0000785; C:chromatin; IEA.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003682; P:chromatin binding; IEA.  
 DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.  
 DR InterPro; IPR000953; Chromo.  
 DR Pfam; PF00385; chromo; 1.  
 DR PRINTS; PRO0504; CHROMODOMAIN.  
 DR SMART; SM00298; CHROMO; 1.  
 DR PROSITE; PS50013; CHROMO 2; 1.  
 SQ SEQUENCE 84 AA; 9934 MW; B4CE22F3261B434F CRC64;  
  
 Query Match 100.0%; Score 28; DB 5; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRPQ 5  
 DB 34 RGRPQ 38  
  
 RESULT 4  
 Q89NY2  
 ID Q89NY2 PRELIMINARY; PRT; 87 AA.  
 AC Q89NY2  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Bsr3701 protein.  
 GN BSR3701  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 ON NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA 110;  
 RX MEDLINE=22484998; PubMed=12597275;  
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
 RA Sasamoto S., Watanabe A., Ideasa K., Iguchi M., Kawashima K.,  
 RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,  
 RA Tabata S.;  
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 RT Bradyrhizobium japonicum USDA110.";  
 RL DNA Res. 9:189-197(2002).  
 DR EMBL; AP005948; BAC48966.1;  
 KW Complete proteome.  
 SQ SEQUENCE 87 AA; 9554 MW; A64BF0F03E2EF0DA CRC64;  
  
 Query Match 100.0%; Score 28; DB 16; Length 87;



Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
DB 82 RGRPQ 86

## RESULT 5

Q52183  
ID Q52183 PRELIMINARY; PRT; 91 AA.  
AC Q52183;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Synchocystis PCC6803 plasmid PCB2.4, complete sequence.  
OS Synchocystis sp. (strain PCC 6803).  
CG Plasmid PCB2.4.  
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PCC 6803;  
RX MEDLINE=94302133; PubMed=8029321;  
RA Yang X., McFadden B.A.;  
RT "The complete DNA sequence and replication analysis of the plasmid  
pCB2.4 from the cyanobacterium Synchocystis PCC 6803."  
RL Plasmid 31:131-137(1994).  
DR EMBL; L25424; AAA97420.1; -.  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
KW Plasmid.  
SQ SEQUENCE 91 AA; 10090 MW; 5695BCA15DB95FEE CRC64;

Query Match 100.0%; Score 28; DB 2; Length 91;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
DB 77 RGRPQ 81

## RESULT 6

Q81ZS3  
ID Q81ZS3 PRELIMINARY; PRT; 122 AA.  
AC Q81ZS3;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE MRD51 protein (Fragment).  
GN MRD51.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Matsumoto M., Weinberger D.R., Straub R.E.;  
RT "Molecular cloning, sequencing, and characterization of a novel 500  
kilobase gene (MRD51) from 6p24, a Schizophrenia candidate region."  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF520803; AAN06678.1; -.  
FT NON TER 1  
SQ SEQUENCE 122 AA; 14063 MW; 720820EDA6C58350 CRC64;

Query Match 100.0%; Score 28; DB 4; Length 122;  
Best Local Similarity 100.0%; Pred. No. 1.2e-02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
DB 107 RGRPQ 111

## RESULT 7

Q7XBG3  
ID Q7XBG3 PRELIMINARY; PRT; 130 AA.  
AC Q7XBG3;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Setaria italica (Fox-rail millet).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC PACCAD clade; Panicoideae; Paniceae; Setaria.  
OX NCBI\_TaxID=4555;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Jing X., Jing Juan Y., Guang Ming A., Qian Z.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY327512; AAP93139.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 130 AA; 14128 MW; F71F045C7329A944 CRC64;

Query Match 100.0%; Score 28; DB 10; Length 130;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
DB 35 RGRPQ 39

## RESULT 8

Q8S1G6  
ID Q8S1G6 PRELIMINARY; PRT; 134 AA.  
AC Q8S1G6;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE P0699H05.19 protein.  
GN P0699H05.19.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoidae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GAS) genomic DNA, chromosome 1, PAC  
clone:P0699H05.19";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP003299; BAA89897.1; -.  
DR Gramene; Q8S1G6; -.  
SQ SEQUENCE 134 AA; 14416 MW; F1C94FDDA92BAD91 CRC64;

Query Match 100.0%; Score 28; DB 10; Length 134;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
DB 81 RGRPQ 85

## RESULT 9

O23967  
ID O23967 PRELIMINARY; PRT; 138 AA.  
AC O23967;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

[1]  
RP SEQUENCE FROM N.A.  
RA Novo F.J.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBSJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RA Oadero M.D.;  
RT "A novel gene in lp36 is disrupted by a translocation in a leukemia patient.";  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AJ310434; CAD27416.1; "-  
SO SEQUENCE 140 AA; 15384 MW; 839BC22236359334 CRC64;  
  
Query Match 100.0%; Score 28; DB 4; Length 140;  
Best Local Similarity 100.0%; Pred. NO. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps  
  
QY 1 RGRPQ 5  
DB 64 RGRPQ 68  
  
RESULT 12  
O48526 PRELIMINARY; PRT; 141 AA.  
ID O48526;  
AC O48526;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE A2942190 protein (Hypothetical protein).  
GN A2942190.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.B., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana";  
RL Nature 402:761-768(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Barth J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,

RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.,  
RT "Arabidopsis Open Reading Frame (ORF) Clones";  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC002561; AAB88644.1; -;  
DR EMBL; AY080869; AAL87340.1; -;  
DR EMBL; AY122991; AAM57524.1; -;  
DR PIR; T00928; T00928.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR000637; AT:hook.  
DR PRINTS; PR00929; ATHOOK.  
KW Hypothetical protein.  
SQ SEQUENCE 141 AA; 16027 MW; 7F450E8BE723C136 CRC64;  
  
Query Match 100.0%; Score 28; DB 10; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 RGRPQ 5  
Db 33 RGRPQ 37  
|||||  
RESULT 13  
Q9M2Q5 PRELIMINARY; PRT; 141 AA.  
AC Q9M2Q5  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN T10K17.140 OR AT3G57930.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
[1]  
RN SEQUENCE FROM N.A.  
RA Benes V., Wurmbach E., Drzonek H., Ansoerge W., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X., Quetier F., Salancubut M.,  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,  
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,  
RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,  
RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.,  
RT "Arabidopsis Full Length cDNA Clones";  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
[4]  
RN SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,  
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,  
RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,  
RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.,  
RT "Arabidopsis Open Reading Frame (ORF) Clones";  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL132977; CAB67621.1; -;  
DR EMBL; BT004228; AAO42243.1; -;  
DR EMBL; BT005705; AAO64125.1; -;

DR PIR; T46015; T46015.  
KW Hypothetical protein.  
SQ SEQUENCE 141 AA; 15940 MW; 518BC0CB6E2D467C CRC64;  
  
Query Match 100.0%; Score 28; DB 10; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 RGRPQ 5  
Db 33 RGRPQ 37  
|||||  
RESULT 14  
Q86NAO PRELIMINARY; PRT; 142 AA.  
AC Q86NAO  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Bric a brac protein (Fragment).  
GN BAB.  
OS Drosophila virilis (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7244;  
[1]  
RN SEQUENCE FROM N.A.  
RA Lours C., Bardot O., Godt D., Laski F., Couderc J.L.;  
RT "The BTB proteins Bric a Brac bind to DNA through a composite DNA-  
binding domain";  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ544785; CAD67617.1; -;  
DR InterPro; IPR007899; HTH\_psq.  
DR Pfam; PF05225; HTH\_psq; 1.  
FT NON\_TER 1  
FT NON\_TER 142  
SQ SEQUENCE 142 AA; 16472 MW; F7B7319C6B260B59 CRC64;  
  
Query Match 100.0%; Score 28; DB 5; Length 142;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 RGRPQ 5  
Db 95 RGRPQ 99  
|||||  
RESULT 15  
Q8LFAS PRELIMINARY; PRT; 142 AA.  
AC Q8LFAS  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
[1]  
RN SEQUENCE FROM N.A.  
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,  
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
annotation";  
RL Genome Biol. 0:0-0(2002).  
[2]  
RN SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.;

RT "Full-length cDNA from Arabidopsis thaliana."  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY084953; AA61514.1; IEA.  
DR GO: GO:0005634; C:nucleus; IEA.  
DR GO: GO:0003677; F:DNA binding; IEA.  
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro: IPR00637; AT hook.  
DR PRINTS: PR0929; ATHOOK.  
KW Hypothetical protein.  
SQ SEQUENCE 142 AA; 16155 MW; B0B5971D20A164E1 CRC64;

Query Match 100.0%; Score 28; DB 10; Length 142;  
Best Local Similarity 100.0%; Pred. NO. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||  
Db 33 RGRPQ 37

Search completed: April 6, 2004, 16:14:37  
Job time : 19.161 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 27.9907 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-1

Perfect score: 28

Sequence: 1 RGRPQ 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04.\*

1: Geneseq1980s.\*

2: Geneseq1990s.\*

3: Geneseq2000s.\*

4: Geneseq2001s.\*

5: Geneseq2002s.\*

6: Geneseq2003as.\*

7: Geneseq2003bs.\*

8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	28	100.0	5	4	AAB48771 Human sal
2	28	100.0	6	4	AAB48772 Human sal
3	28	100.0	7	4	AAB48773 Human sal
4	28	100.0	8	4	AAB48774 Human sal
5	28	100.0	9	4	AAB48775 Human sal
6	28	100.0	10	4	AAB48776 Human sal
7	28	100.0	16	6	ABP82118 G protein
8	28	100.0	12	4	AAB48783 Human sal
9	28	100.0	13	5	ABP78083 Extracell
10	28	100.0	22	4	AAM19718 Peptide #
11	28	100.0	22	4	ABP39457 Peptide #
12	28	100.0	22	4	AAM32988 Peptide #
13	28	100.0	22	4	ABP24221 Protein #
14	28	100.0	22	4	AAM72760 Human bon
15	28	100.0	22	4	AAM60145 Human bon
16	28	100.0	22	4	ABG54459 Human liv
17	28	100.0	22	5	ABG42584 Human pep
18	28	100.0	37	2	AAR60487 Peptide #
19	28	100.0	37	4	AAM14754 Peptide #
20	28	100.0	37	4	ABP33719 Peptide #
21	28	100.0	37	4	AAM27177 Peptide #
22	28	100.0	37	4	ABP28535 Peptide #
23	28	100.0	37	4	ABP3166 Protein #
24	28	100.0	37	4	AAM66893 Human bon
25	28	100.0	37	4	AAM54488 Human bra

26	28	100.0	37	4	ABG48557 Human liv
27	28	100.0	37	4	AAM02477 Peptide #
28	28	100.0	37	5	ABG36549 Human pep
29	28	100.0	51	4	AAM95786 Human rep
30	28	100.0	51	4	ABP96317 Human tes
31	28	100.0	52	5	ABP64175 Human ORF
32	28	100.0	55	4	AU63381 Propionib
33	28	100.0	55	5	ABP35444 Human ORF
34	28	100.0	55	5	ABM59900 Propionib
35	28	100.0	56	4	AU64411 Propionib
36	28	100.0	56	6	ABM60930 Propionib
37	28	100.0	74	1	AAP80105 Sequence
38	28	100.0	74	1	AAP91656 Clone MC1
39	28	100.0	78	4	AA00192 Human pol
40	28	100.0	78	7	ADC95875 E. faeciu
41	28	100.0	84	4	ABB66774 Drosophil
42	28	100.0	87	2	AY04818 Mycobacte
43	28	100.0	90	4	ABG17940 Novel hum
44	28	100.0	90	6	ABO0927 Polypepti
45	28	100.0	95	4	ABG09699 Novel hum

ALIGNMENTS

RESULT 1  
AAB48771  
ID AAB48771 standard; peptide; 5 AA.

AC AAB48771;

XX 09-MAR-2001 (first entry)

DE Human saliva PRP-1 fragment (residues 106-110), SEQ ID NO:1.

KW Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

PT New oligopeptides comprising 2 arginine residues from degradation of

PT proline-rich proteins, useful for preventing dental caries.

XX Claim 3; Page 24; 36pp; English.

The invention relates to human PRP-1-derived oligopeptides (AAB48771-AAB48783) which contain at least two arginine residues and which protect against dental caries. PRPs (proline-rich proteins) are salivary proteins encoded by six clustered genes on chromosome 12p13.2 and are potential determinants of a person's susceptibility to dental caries. PRPs are degraded by Actinomyces and Streptococcus species to small peptide fragments. These are metabolised by oral bacteria for nutritional purposes, with certain bacterial species generating ammonia via the catabolism of arginine. The peptides of the invention, being arginine-rich, can also be converted to ammonia by these bacteria. The ammonia thus formed raises the pH at the dental surface, thereby protecting the teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

Applicant

CC derived oligopeptides of the invention

XX Sequence 5 AA;

Query Match 100.0%; Score 28; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RGRPQ 5  
|||  
Db 1 RGRPQ 5

## RESULT 2

AAB48772  
ID AAB48772 standard; peptide; 6 AA.

XX AAB48772;

AC AAB48772;

XX 09-MAR-2001 (first entry)

DT Human saliva PRP-1 fragment (residues 106-111), SEQ ID NO:2.

DE Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX Homo sapiens.

OS WO200069890-A1.

PN 23-NOV-2000.

PD 11-MAY-2000; 2000WO-SE000930.

PF 17-MAY-1999; 99SE-00001773.

PR (STRO/) STROEMBERG N.  
(JOHA/) JOHANSSON I.

PA Stroemberg N, Johansson I;

PI WPI; 2001-031923/04.

DR New oligopeptides comprising 2 arginine residues from degradation of

PT proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

The invention relates to human PRP-1-derived oligopeptides (AAB48771-AAB48783) which contain at least two arginine residues and which protect against dental caries. PRPs (proline-rich proteins) are salivary proteins encoded by six clustered genes on chromosome 12p13.2 and are potential determinants of a person's susceptibility to dental caries. PRPs are degraded by Actinomyces and Streptococcus species to small peptide fragments. These are metabolised by oral bacteria for nutritional purposes, with certain bacterial species generating ammonia via the catabolism of arginine. The peptides of the invention, being arginine-rich, can also be converted to ammonia by these bacteria. The ammonia thus formed raises the pH at the dental surface, thereby protecting the teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-derived oligopeptides of the invention

XX Sequence 6 AA;

Query Match 100.0%; Score 28; DB 4; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RGRPQ 5  
|||  
Db 1 RGRPQ 5

## RESULT 3

AAB48773  
ID AAB48773 standard; peptide; 7 AA.

XX AAB48773;

AC 09-MAR-2001 (first entry)

DT Human saliva PRP-1 fragment (residues 106-112), SEQ ID NO:3.

DE Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX Homo sapiens.

OS WO200069890-A1.

PN 23-NOV-2000.

PD 11-MAY-2000; 2000WO-SE000930.

PF 17-MAY-1999; 99SE-00001773.

PR (STRO/) STROEMBERG N.  
(JOHA/) JOHANSSON I.

PA Stroemberg N, Johansson I;

PI WPI; 2001-031923/04.

DR New oligopeptides comprising 2 arginine residues from degradation of

PT proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

The invention relates to human PRP-1-derived oligopeptides (AAB48771-AAB48783) which contain at least two arginine residues and which protect against dental caries. PRPs (proline-rich proteins) are salivary proteins encoded by six clustered genes on chromosome 12p13.2 and are potential determinants of a person's susceptibility to dental caries. PRPs are degraded by Actinomyces and Streptococcus species to small peptide fragments. These are metabolised by oral bacteria for nutritional purposes, with certain bacterial species generating ammonia via the catabolism of arginine. The peptides of the invention, being arginine-rich, can also be converted to ammonia by these bacteria. The ammonia thus formed raises the pH at the dental surface, thereby protecting the teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-derived oligopeptides of the invention

XX Sequence 7 AA;

Query Match 100.0%; Score 28; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RGRPQ 5  
|||  
Db 1 RGRPQ 5

## RESULT 4

AAB48774  
ID AAB48774 standard; peptide; 8 AA.

XX AAB48774;

AC 09-MAR-2001 (first entry)

DT Human saliva PRP-1 fragment (residues 106-113), SEQ ID NO:4.

DE Human; PRP-1; proline-rich protein; saliva; dental caries;

XX

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.

XX Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention

XX Sequence 8 AA;

Query Match 100.0%; Score 28; DB 4; Length 8;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RGRPQ 5

Db 1 RGRPQ 5

RESULT 5

AAB48775

ID AAB48775 standard; peptide; 9 AA.

XX AC AAB48775;

XX 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 106-114), SEQ ID NO:5.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX Stroemberg N, Johansson I;  
 XX WPI; 2001-031923/04.  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention

XX Sequence 9 AA;

Query Match 100.0%; Score 28; DB 4; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RGRPQ 5

Db 1 RGRPQ 5

RESULT 6

AAB48776

ID AAB48776 standard; peptide; 10 AA.

XX AC AAB48776;

XX 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 106-115), SEQ ID NO:6.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.





CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences ABA48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention

XX SQ Sequence 17 AA;

Query Match 100.0%; Score 28; DB 4; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
 |||||  
 DB 8 RGRPQ 12

RESULT 9  
 ABB78083  
 ID ABB78083 standard; peptide; 19 AA.

AC ABB78083;  
 DT 05-NOV-2002 (first entry)

DE Extracellular loop, comprising an epitope, from alpha-2B receptor.

XX Targeting component; alpha-2B adrenergic receptor; ribosome inactivation;  
 KW alpha-C adrenergic receptor; pain; chronic pain; visceral pain;  
 KW neuropathic pain; referred pain; allodynia type pain; cancer;  
 KW irritable bowel syndrome; extracellular loop; epitope.

XX Unidentified.

OS WO200253177-A2.

PN 11-JUL-2002.

PD 14-DEC-2001; 2001WO-US048651.

PF 29-DEC-2000; 2000US-00751053.

PR (ALLR ) ALLERGAN SALES INC.

PA Gil DW, Aoki KR;

PI WPI; 2002-619081/66.

PS Agent for treating pain such as neuropathic pain comprises a therapeutic  
 component and a targeting component.

XX Claim 29; Page 68; 76pp; English.

CC The specification describes an agent, which comprises a therapeutic  
 CC component and a targeting component, where the targeting component  
 CC selectively binds at the alpha-2B/alpha-C adrenergic receptor subtype as  
 CC compared to the alpha-2A adrenergic receptor subtype. The therapeutic  
 CC component inactivates cellular ribosomes. The therapeutic agent is used  
 CC for treating pain such as chronic pain, visceral pain, neuropathic pain,  
 CC referred pain and allodynia type pain (persisting from 2-27 months)  
 CC without affecting acute pain sensation or tactile sensation such as  
 CC chronic pain, visceral pain, neuropathic pain, referred pain and  
 CC allodynia type pain and for treating pain associated with cancer and  
 CC irritable bowel syndrome. The present sequence represents an  
 CC extracellular loop, comprising an epitope, from the alpha-2B receptor. It  
 CC is used to raise antibodies, which can then form the targeting component  
 CC of the agent of the invention

XX SQ Sequence 19 AA;

Query Match 100.0%; Score 28; DB 5; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
 |||||  
 DB 9 RGRPQ 13

RESULT 10

AAM19718

ID AAM19718 standard; protein; 22 AA.

AC AAM19718;

DT 12-OCT-2001 (first entry)

DE Peptide #6152 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000670.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00832366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 gene expression in human cervical epithelial cells.

XX Claim 27; SEQ ID NO 24544; 487pp; English.

CC The present invention relates to human single exon nucleic acid probes  
 CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 22 AA;

Query Match 100.0%; Score 28; DB 4; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
 |||||  
 DB 12 RGRPQ 16

```
RESULT 11
ABB39457
ID ABB39457 standard; peptide; 22 AA.
XX
AC ABB39457;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #6963 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
FN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human foetal liver.
XX
PS Claim 27; SEQ ID NO 32092; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC single gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 22 AA;
XX
Query Match 100.0%; Score 28; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 RGRPQ 5
DB 12 RGRPQ 16
|||||
RESULT 12
AAM32988
ID AAM32988 standard; protein; 22 AA.
XX
AC AAM32988;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #7025 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX
OS genetic disorder.
XX
```

```
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
XX
PS Claim 27; SEQ ID NO 33257; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
CC see AA131315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX
SQ Sequence 22 AA;
XX
Query Match 100.0%; Score 28; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 RGRPQ 5
DB 12 RGRPQ 16
|||||
RESULT 13
ABB24221
ID ABB24221 standard; protein; 22 AA.
XX
AC ABB24221;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #8220 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000666.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
```

PR 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-48899/53.  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 XX hearts.  
 PT  
 XX  
 XX Claim 15; SEQ ID NO 25991; 530pp; English.  
 XX  
 XX The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC AB21535-AB41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,  
 CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 22 AA;  
 SQ

Query Match 100.0%; Score 28; DB 4; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
 Db 12 RGRPQ 16  
 |||||

RESULT 14  
 AAM72760  
 ID AAM72760 standard; protein; 22 AA.  
 XX  
 XX AAM72760;  
 AC  
 XX  
 XX 06-NOV-2001 (first entry)  
 DT  
 XX Human bone marrow expressed probe encoded protein SEQ ID NO: 33056.  
 DE  
 XX Human; bone marrow expressed exon; gene expression analysis; probe;  
 XX microarray; cancer; leukaemia; lymphoma; myeloma.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200157276-A2.  
 XX  
 XX 09-AUG-2001.  
 PD  
 XX  
 XX 30-JAN-2001; 2001WO-US000668.  
 PF  
 XX  
 XX 04-FEB-2000; 2000US-0180312P.  
 PR  
 XX 26-MAY-2000; 2000US-0207456P.  
 PR  
 XX 30-JUN-2000; 2000US-00608408.  
 PR  
 XX 03-AUG-2000; 2000US-00632366.  
 PR  
 XX 21-SEP-2000; 2000US-0234687P.  
 PR  
 XX 27-SEP-2000; 2000US-0236359P.  
 PR  
 XX 04-OCT-2000; 2000GB-00024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-48899/53.  
 XX  
 XX Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human bone marrow.  
 XX  
 XX Example 4; SEQ ID NO 33066; 658pp + Sequence Listing; English.  
 PS  
 XX  
 XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention  
 XX  
 XX Sequence 22 AA;  
 SQ

Query Match 100.0%; Score 28; DB 4; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
 Db 12 RGRPQ 16  
 |||||

RESULT 15  
 AAM60146  
 ID AAM60146 standard; protein; 22 AA.  
 XX  
 XX AAM60146;  
 AC  
 XX  
 XX 05-NOV-2001 (first entry)  
 DT  
 XX Human brain expressed single exon probe encoded protein SEQ ID NO: 32251.  
 DE  
 XX Human; brain expressed exon; gene expression analysis; probe; microarray;  
 XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200157275-A2.  
 XX  
 XX 09-AUG-2001.  
 PD  
 XX  
 XX 30-JAN-2001; 2001WO-US000667.  
 PF  
 XX  
 XX 04-FEB-2000; 2000US-0180312P.  
 PR  
 XX 26-MAY-2000; 2000US-0207456P.  
 PR  
 XX 30-JUN-2000; 2000US-00608408.  
 PR  
 XX 03-AUG-2000; 2000US-00632366.  
 PR  
 XX 21-SEP-2000; 2000US-0234687P.  
 PR  
 XX 27-SEP-2000; 2000US-0236359P.  
 PR  
 XX 04-OCT-2000; 2000GB-00024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-483446/52.  
 XX  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 XX brains.  
 PT  
 XX  
 XX Example 4; SEQ ID NO 32251; 650pp + Sequence Listing; English.  
 PS  
 XX  
 XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention  
 XX  
 XX Sequence 22 AA;  
 SQ

Query Match 100.0%; Score 28; DB 4; Length 22;

Best Local Similarity 100.0%; Pred. No. 57;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RGRPQ 5  
Db 12 RGRPQ 16

Search completed: April 6, 2004, 16:06:35  
Job time : 28.9907 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 19.6262 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-1

Perfect score: 28

Sequence: 1 RGRPQ 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	16	14	US-10-225-567A-791
2	28	100.0	22	9	US-09-864-761-39519
3	28	100.0	37	9	US-09-864-761-34464
4	28	100.0	42	12	US-10-424-599-163763
5	28	100.0	51	10	US-09-764-891-4444
6	28	100.0	52	9	US-09-867-550-1090
7	28	100.0	55	11	US-09-864-408A-8834
8	28	100.0	79	12	US-10-424-599-177907
9	28	100.0	83	12	US-10-424-599-183421
10	28	100.0	89	12	US-10-424-599-195980
11	28	100.0	100	12	US-10-425-114-60575
12	28	100.0	108	12	US-10-424-599-261482
13	28	100.0	111	12	US-10-424-599-245383
14	28	100.0	123	12	US-10-424-599-232489
15	28	100.0	132	9	US-09-864-761-43644

16 28 100.0 135 12 US-10-425-114-44223 Sequence 44223, A  
17 28 100.0 137 12 US-10-424-599-246391 Sequence 246391, A  
18 28 100.0 143 12 US-10-424-599-214108 Sequence 214108, A  
19 28 100.0 163 12 US-10-424-599-189117 Sequence 189117, A  
20 28 100.0 166 14 US-10-157-031-80 Sequence 80, Appl  
21 28 100.0 183 12 US-10-425-114-70816 Sequence 70816, A  
22 28 100.0 188 12 US-10-425-114-49852 Sequence 49852, A  
23 28 100.0 199 12 US-10-425-114-60578 Sequence 60578, A  
24 28 100.0 209 15 US-10-108-360A-4370 Sequence 4370, Ap  
25 28 100.0 220 12 US-10-425-114-41747 Sequence 41747, A  
26 28 100.0 267 14 US-10-156-761-14846 Sequence 14846, A  
27 28 100.0 268 12 US-10-425-114-59678 Sequence 59678, A  
28 28 100.0 270 12 US-10-282-122A-60255 Sequence 60255, A  
29 28 100.0 289 15 US-10-369-493-22900 Sequence 22900, A  
30 28 100.0 300 12 US-10-425-114-64837 Sequence 64837, A  
31 28 100.0 301 12 US-10-425-114-66877 Sequence 46877, A  
32 28 100.0 301 15 US-10-369-493-518 Sequence 518, App  
33 28 100.0 305 9 US-09-738-626-6463 Sequence 6463, Ap  
34 28 100.0 311 12 US-10-425-114-45282 Sequence 45282, A  
35 28 100.0 311 12 US-10-425-114-59180 Sequence 59180, A  
36 28 100.0 311 12 US-10-425-114-61343 Sequence 61343, A  
37 28 100.0 318 12 US-10-425-114-37514 Sequence 37514, A  
38 28 100.0 318 12 US-10-425-114-50661 Sequence 50661, A  
39 28 100.0 329 12 US-10-425-114-57144 Sequence 57144, A  
40 28 100.0 332 12 US-10-425-114-59820 Sequence 59820, A  
41 28 100.0 336 12 US-10-425-114-61384 Sequence 61384, A  
42 28 100.0 339 12 US-10-425-114-45434 Sequence 45434, A  
43 28 100.0 350 14 US-10-156-761-7775 Sequence 7775, Ap  
44 28 100.0 352 12 US-10-282-122A-59067 Sequence 59067, A  
45 28 100.0 403 12 US-10-425-114-41952 Sequence 41952, A

#### ALIGNMENTS

##### RESULT 1

US-10-225-567A-791

; Sequence 791, Application US/10225567A

; Publication No. US20030113798A1

; GENERAL INFORMATION:

; APPLICANT: LifeSpan Biosciences

; APPLICANT: Brown, Joseph P.

; APPLICANT: Burmer, Glenna C.

; APPLICANT: Roush, Christine L.

; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTOR

; FILE REFERENCE: 1920-4-4

; CURRENT APPLICATION NUMBER: US/10/225,567A

; CURRENT FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/257,144

; PRIOR FILING DATE: 2000-12-19

; NUMBER OF SEQ ID NOS: 2392

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 791

; LENGTH: 16

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-225-567A-791

Query Match 100.0%; Score 28; DB 14; Length 16;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5

DB 11 RGRPQ 15

##### RESULT 2

US-09-864-761-39519

; Sequence 39519, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

```
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39519
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004527.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
US-09-864-761-39519

Query Match 100.0%; Score 28; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5
Db 12 RGRPQ 16

RESULT 3
US-09-864-761-34464
```

```
; Sequence 34464, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,587
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34464
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007179.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.85
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.2
; OTHER INFORMATION: EST HUMAN HIT: BF310193.1, EVALUATE 7.40e+00
; OTHER INFORMATION: SWISSPROT HIT: P21573, EVALUATE 4.70e+00
US-09-864-761-34464

Query Match 100.0%; Score 28; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



QY 1 RGRPQ 5  
|  
|  
|  
|  
Db 13 RGRPQ 17

## RESULT 4

US-10-424-599-163763  
; Sequence 163763, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 163763  
; LENGTH: 42  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_118898C.1.psp  
US-10-424-599-163763

Query Match 100.0%; Score 28; DB 12; Length 42;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|  
|  
|  
|  
Db 20 RGRPQ 24

## RESULT 5

US-09-764-891-4444  
; Sequence 4444, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4444  
; LENGTH: 51  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (43)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-4444

Query Match 100.0%; Score 28; DB 10; Length 51;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|  
|  
|  
|  
Db 2 RGRPQ 6

## RESULT 6

US-09-867-550-1090  
; Sequence 1090, Application US/09867550

Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells an  
; TITLE OF INVENTION: Thereby  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1090  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-867-550-1090

Query Match 100.0%; Score 28; DB 9; Length 52;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|  
|  
|  
|  
Db 8 RGRPQ 12

## RESULT 7

US-09-864-408A-8834  
; Sequence 8834, Application US/09864408A  
; Publication No. US20040009474A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Shimkets, Richard A.  
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides En  
; FILE REFERENCE: 21402-012  
; CURRENT APPLICATION NUMBER: US/09/864,408A  
; CURRENT FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: 60/206,690  
; PRIOR FILING DATE: 2000-05-24  
; NUMBER OF SEQ ID NOS: 9068  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8834  
; LENGTH: 55  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-864-408A-8834

Query Match 100.0%; Score 28; DB 11; Length 55;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|  
|  
|  
|  
Db 34 RGRPQ 38

## RESULT 8

US-10-424-599-177907  
; Sequence 177907, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 177907  
; LENGTH: 79  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_131666C.1.pap  
US-10-424-599-177907

Query Match 100.0%; Score 28; DB 12; Length 79;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||  
Db 73 RGRPQ 77

RESULT 9  
US-10-424-599-183421  
; Sequence 183421, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J.  
; APPLICANT: Kovalic David K.  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 183421  
; LENGTH: 83  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)-(83)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_136642C.1.pap  
US-10-424-599-183421

Query Match 100.0%; Score 28; DB 12; Length 83;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||  
Db 5 RGRPQ 9

RESULT 10  
US-10-424-599-195980  
; Sequence 195980, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J.  
; APPLICANT: Kovalic David K.  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 195980  
; LENGTH: 89

; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_18997C.1.pap  
US-10-424-599-195980

Query Match 100.0%; Score 28; DB 12; Length 89;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||  
Db 22 RGRPQ 26

RESULT 11  
US-10-425-114-60975  
; Sequence 60975, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 60975  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3356-034-GS\_FLI.pap  
US-10-425-114-60975

Query Match 100.0%; Score 28; DB 12; Length 100;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||  
Db 59 RGRPQ 63

RESULT 12  
US-10-424-599-261482  
; Sequence 261482, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J.  
; APPLICANT: Kovalic David K.  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 261482  
; LENGTH: 108  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_78140C.1.pap  
US-10-424-599-261482

Query Match 100.0%; Score 28; DB 12; Length 108;

Best Local Similarity 100.0%; Pred. No. 2.8e+02; DB 12; Length 111;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
DB 1 RGRPQ 5

RESULT 13  
US-10-424-599-245383  
; Sequence 245383, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 245383  
; LENGTH: 111  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_63611C.1.pep  
US-10-424-599-245383

Query Match 100.0%; Score 28; DB 12; Length 111;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
DB 107 RGRPQ 111

RESULT 14  
US-10-424-599-232489  
; Sequence 232489, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 232489  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_51961C.1.pep  
US-10-424-599-232489

Query Match 100.0%; Score 28; DB 12; Length 123;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
DB 33 RGRPQ 37

RESULT 15

US-09-864-761-43644  
; Sequence 43644, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 43644  
; LENGTH: 132  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006518.17  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77  
; OTHER INFORMATION: EST HUMAN HIT: BF088785.1, EVALUE 1.00e-01  
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02  
US-09-864-761-43644

Query Match 100.0%; Score 28; DB 9; Length 132;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
|||||

D> 88 RGRPQ 92

Search completed: April 6, 2004, 17:05:50  
Job time : 21.6262 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 7.33645 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-1

Perfect score: 28

Sequence: 1 RGRFQ 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgm2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgm2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgm2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgm2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgm2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /cgm2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	78	4	US-09-107-532A-5502
2	28	100.0	95	2	US-08-341-843B-4
3	28	100.0	95	2	US-08-457-497E-9
4	28	100.0	118	4	US-09-489-039A-8588
5	28	100.0	176	4	US-09-252-991A-30542
6	28	100.0	225	4	US-09-489-039A-11152
7	28	100.0	230	4	US-09-252-991A-16763
8	28	100.0	271	4	US-09-252-991A-26022
9	28	100.0	296	4	US-09-252-991A-17410
10	28	100.0	300	4	US-09-252-991A-27253
11	28	100.0	308	4	US-09-252-991A-29243
12	28	100.0	328	4	US-09-252-991A-27822
13	28	100.0	329	4	US-09-489-039A-11151
14	28	100.0	330	1	US-08-118-270-20
15	28	100.0	330	5	PCT-US93-08528-20
16	28	100.0	331	4	US-09-252-991A-28389
17	28	100.0	372	4	US-09-252-991A-19721
18	28	100.0	389	4	US-09-252-991A-30972
19	28	100.0	392	4	US-09-252-991A-27798
20	28	100.0	428	4	US-09-252-991A-25955
21	28	100.0	468	4	US-09-252-991A-17753
22	28	100.0	477	4	US-09-252-991A-19922
23	28	100.0	508	4	US-09-252-991A-32651
24	28	100.0	519	4	US-09-252-991A-19734
25	28	100.0	600	5	PCT-US95-10166-2
26	28	100.0	651	4	US-09-252-991A-32204
27	28	100.0	667	4	US-09-252-991A-26436

28	28	100.0	675	4	US-09-252-991A-23662	Sequence 23662, A
29	28	100.0	688	4	US-09-252-991A-21780	Sequence 21780, A
30	28	100.0	690	4	US-09-252-991A-24903	Sequence 24903, A
31	28	100.0	726	4	US-09-252-991A-20675	Sequence 20675, A
32	28	100.0	748	4	US-09-252-991A-18427	Sequence 18427, A
33	28	100.0	1008	4	US-09-252-991A-29419	Sequence 29419, A
34	28	100.0	1151	4	US-09-252-991A-21328	Sequence 21328, A
35	28	100.0	1253	4	US-09-489-039A-12097	Sequence 12097, A
36	28	100.0	1260	4	US-08-506-296B-21	Sequence 21, Appl
37	28	100.0	1282	4	US-09-543-681A-5419	Sequence 5419, Ap
38	28	100.0	1646	4	US-09-252-991A-22312	Sequence 22312, A
39	28	100.0	1673	4	US-09-418-710-70	Sequence 70, Appl
40	28	100.0	1674	4	US-09-418-710-1	Sequence 1, Appl
41	25	89.3	37	2	US-08-598-873-17	Sequence 17, Appl
42	25	89.3	37	3	US-08-605-430-17	Sequence 17, Appl
43	25	89.3	133	4	US-09-252-991A-18533	Sequence 18533, A
44	25	89.3	140	4	US-09-252-991A-31623	Sequence 31623, A
45	25	89.3	158	3	US-08-851-843A-82	Sequence 82, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-107-532A-5502  
; Sequence 5502, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5502:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:

NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...78  
SEQUENCE DESCRIPTION: SEQ ID NO: 5502:  
US-09-107-532A-5502

Query Match 100.0%; Score 28; DB 4; Length 78;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
Db 65 RGRPQ 69

## RESULT 2

US-08-341-843B-4  
; Sequence 4, Application US/08341843B  
; Patent No. 5872225  
; GENERAL INFORMATION:  
; APPLICANT: Lemmon, Vance  
; TITLE OF INVENTION: A Method for Characterizing the  
; TITLE OF INVENTION: Nucleotide Sequence of LICAM and  
; Patent No. 5872225  
; TITLE OF INVENTION: the Nucleotide Sequence  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,  
; ADDRESSEE: Minnich & McKee  
; STREET: 1100 Superior Avenue  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: U.S.A.  
; ZIP: 44114-2518  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb  
; MEDIUM TYPE: storable  
; COMPUTER: Compaq Prolinea 5100e  
; OPERATING SYSTEM: DOS 5.0  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/341,843B  
; FILING DATE: No. 5872225ember 18, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/904,991  
; FILING DATE: June 26, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Minnich, Richard J.  
; REGISTRATION NUMBER: 24,175  
; REFERENCE/DOCKET NUMBER: CWR 2 149-1  
; TELEPHONE: (216) 861-5582  
; TELEFAX: (216) 241-1666  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: amino acids  
; HYPOTHETICAL: irrelevant  
; ANTI-SENSE: no  
; ORIGINAL SOURCE:  
; ORGANISM: mouse  
; INDIVIDUAL ISOLATE: 8 day old mouse brain  
; IMMEDIATE SOURCE:  
; LIBRARY: lamda GT 10 and lamda GT11  
; CLONE: synthesis of several clones  
; PUBLICATION INFORMATION:  
; AUTHORS: Moos, M.  
; AUTHORS: Tacke, R.  
; AUTHORS: Scherer, H.  
; AUTHORS: Teplov, D.  
; AUTHORS: Fruh, K.

; AUTHORS: Schachner, M.  
; TITLE: Neural adhesion molecule L1 is a  
; TITLE: member of the immunoglobulin  
; TITLE: superfamily with binding domains  
; TITLE: similar to fibronectin  
; JOURNAL: NATURE  
; VOLUME: 334  
; ISSUE:  
; PAGES: 701-703  
; DATE: 1988  
; US-08-341-843B-4

Query Match 100.0%; Score 28; DB 2; Length 95;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
Db 25 RGRPQ 29

## RESULT 3

US-08-427-497E-9  
; Sequence 9, Application US/08427497E  
; Patent No. 5969124  
; GENERAL INFORMATION:  
; APPLICANT: Lemmon, Vance  
; TITLE OF INVENTION: A Method for Characterizing the  
; TITLE OF INVENTION: Nucleotide Sequence of LICAM and  
; Patent No. 5969124  
; TITLE OF INVENTION: the Nucleotide Sequence  
; TITLE OF INVENTION: Characterized Thereby  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,  
; ADDRESSEE: Minnich & McKee  
; STREET: 1100 Superior Avenue  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: U.S.A.  
; ZIP: 44114-2518  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb  
; MEDIUM TYPE: storable  
; COMPUTER: Compaq Prolinea 5100e  
; OPERATING SYSTEM: DOS 5.0  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/427,497E  
; FILING DATE: April 24, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/904,991  
; FILING DATE: June 26, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Minnich, Richard J.  
; REGISTRATION NUMBER: 24,175  
; REFERENCE/DOCKET NUMBER: CWR 2 149-3-1  
; TELEPHONE: (216) 861-5582  
; TELEFAX: (216) 241-1666  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: amino acids  
; HYPOTHETICAL: irrelevant  
; ANTI-SENSE: no  
; ORIGINAL SOURCE:

CURRENT FILING DATE: 1999-02-19  
PRIOR APPLICATION NUMBER: US 60/074,788

```
Matches 5; Conservative 0; Mismatches 0; Indels
```

QY 1 RGRPQ 5  
Db 74 RGRPQ 78

## RESULT 8

US-09-252-991A-26022  
; Sequence 26022, Application US/09252991A  
; Patent No. 6551795

## ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26022  
; LENGTH: 271  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26022

Query Match 100.0%; Score 28; DB 4; Length 271;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
Db 98 RGRPQ 102

## RESULT 9

US-09-252-991A-17410  
; Sequence 17410, Application US/09252991A  
; Patent No. 6551795

## ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 17410

; LENGTH: 296

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-17410

Query Match 100.0%; Score 28; DB 4; Length 296;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
Db 16 RGRPQ 20

## RESULT 10

US-09-252-991A-27253  
; Sequence 27253, Application US/09252991A  
; Patent No. 6551795

## ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27253  
; LENGTH: 300  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27253

Query Match 100.0%; Score 28; DB 4; Length 300;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
Db 50 RGRPQ 54

## RESULT 11

US-09-252-991A-29243  
; Sequence 29243, Application US/09252991A  
; Patent No. 6551795

## ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29243  
; LENGTH: 308  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29243

Query Match 100.0%; Score 28; DB 4; Length 308;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQ 5  
Db 147 RGRPQ 151

## RESULT 12

US-09-252-991A-27822  
; Sequence 27822, Application US/09252991A  
; Patent No. 6551795

## ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142



; SEQ ID NO 27822  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27822

Query Match 100.0%; Score 28; DB 4; Length 328;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
Db 45 RGRPQ 49

## RESULT 13

US-09-489-039A-11151  
; Sequence 11151, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 11151  
; LENGTH: 329  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-11151

Query Match 100.0%; Score 28; DB 4; Length 329;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
Db 67 RGRPQ 71

## RESULT 14

US-08-118-270-20  
; Sequence 20, Application US/08118270  
; Patent No. 5508364  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Randall B.  
; APPLICANT: Schuster, David I.  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/118,270  
; FILING DATE: 09-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/943,236  
; FILING DATE: 10-SEP-1992  
; ATTORNEY/AGENT INFORMATION:

; NAME: Townsend, Kevin G.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: MURPHY-2A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633

; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 330 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-118-270-20

Query Match 100.0%; Score 28; DB 1; Length 330;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5  
Db 148 RGRPQ 152

## RESULT 15

PCT-US93-08528-20  
; Sequence 20, Application PC/TUS9308528  
; GENERAL INFORMATION:  
; APPLICANT: New York University  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/08528  
; FILING DATE: 09-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/943,236  
; FILING DATE: 10-SEP-1992  
; ATTORNEY/AGENT INFORMATION:

; NAME: Townsend, Kevin G.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: MURPHY-2 PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633

; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 330 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
PCT-US93-08528-20

Query Match 100.0%; Score 28; DB 5; Length 330;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQ 5

Db       |||||  
         148 RCRPQ 152

Search completed: April 6, 2004, 16:19:38  
Job time : 8.33645 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 6.72897 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-2

Perfect score: 34

Sequence: 1 RGRPQG 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	166	PIHUSC	salivary proline-r
2	34	100.0	166	B25372	salivary proline-r
3	34	100.0	171	A27307	proline-rich phosph
4	34	100.0	2188	A70984	probable polyketid
5	31	91.2	172	1 NKVQBC	core protein p19 -
6	31	91.2	195	C95286	probable transposa
7	31	91.2	284	S76077	hypothetical prote
8	31	91.2	1030	S11034	gene P1 protein -
9	31	91.2	3591	S21010	filamentous hemagg
10	30	88.2	90	JC5931	high mobility grou
11	30	88.2	95	A31895	nonhistone chromos
12	30	88.2	96	JC5933	high mobility grou
13	30	88.2	105	JC5932	high mobility grou
14	30	88.2	107	A32794	nonhistone chromos
15	30	88.2	108	JC4575	high mobility grou
16	30	88.2	109	JC2232	high mobility grou
17	30	88.2	123	H72709	hypothetical prote
18	30	88.2	188	D29149	proline-rich prote
19	30	88.2	191	T28682	hypothetical prote
20	30	88.2	206	G86279	Flu117.27 protein
21	30	88.2	210	T41982	hypothetical prote
22	30	88.2	218	1 TVHURR	transforming prote
23	30	88.2	237	T13649	hypothetical prote
24	30	88.2	255	T46350	hypothetical prote
25	30	88.2	264	T51304	splicing factor RS
26	30	88.2	264	E72642	hypothetical prote
27	30	88.2	271	T47978	splicing factor RS
28	30	88.2	302	T01540	hypothetical prote
29	30	88.2	317	A28996	proline-rich prote

#### RESULT 1

##### PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N;Alternate names: salivary acidic proline-rich protein PRH2

C;Species: Homo sapiens (man)

C;Date: 31-Mar-1981 #sequence revision 12-Apr-1996 #text change 08-Dec-2000

C;Accession: A25372; A19803; B57868; A92277; A92254; A94425; A91954; S02564; S02563; J

R;Maeda, N.; Kim, H.S.; Aren, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A;Title: Differential RNA splicing and post-translational cleavages in the human saliv

A;Reference number: A92492; MUID:85289325; PMID:2993301

A;Accession: A25372

A;Molecule type: mRNA

A;Residues: 1-166 <NAE>

A;Cross-references: GB:K03202; NID:G190481; PIDN:AAA60183.1; PID:G190482

R;Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A;Title: Primary structure of the active tryptic fragments of human and monkey salivar

A;Reference number: A91757; MUID:81191179; PMID:7228490

A;Accession: A19803

A;Molecule type: protein

A;Residues: 17-46 <SCH>

R;Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A;Title: Structures of two HaeIII-type genes in the human salivary proline-rich protei

A;Reference number: A57868; MUID:86196106; PMID:3009472

A;Accession: B57868

A;Molecule type: DNA

A;Residues: 1-166 <KIM>

A;Cross-references: GB:M13058; NID:G190513; PIDN:AAA98808.1; PID:G190514

R;Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A;Title: The primary structure of a salivary calcium-binding proline-rich phosphoprote

A;Reference number: A92277; MUID:80204368; PMID:7380845

A;Contents: protein C

A;Accession: A92277

A;Molecule type: protein

A;Residues: 17-19, 'N', 21-166 <WON>

A;Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite for

R;Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A;Title: The complete primary structure of a proline-rich phosphoprotein from human sa

A;Reference number: A92254; MUID:79173237; PMID:438215

A;Contents: protein A

A;Accession: A92254

A;Molecule type: protein

A;Residues: 17-19, 'N', 21-122 <WO2>

R;Schlesinger, D.H.; Hay, D.I.

In peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gros

A;Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A;Reference number: A94425

salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human

C/Species: Homo sapiens (man)  
 C/Date: 29-Aug-1987 #sequence revision 29-Aug-1987 #text change 20-Aug-1999  
 C/Accession: B25372; A57868; S02562; G38355; S06153; B27307  
 R/Isomura, S.; Saitoh, E.; Sanada, K.  
 J. Biochem. 87, 1071-1077, 1980  
 A/Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relationship to the proline-rich phosphoprotein precursor PRH1 (allele PIF)  
 A/Reference number: A91954; MUID:80227634; PMID:77390979  
 A/Contents: Peptide P-C  
 A/Accession: A91954  
 A/Molecule type: protein  
 A/Residues: 123-166 <ISE>  
 R/Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A/Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A/Reference number: S02562; MUID:89061650; PMID:3196309  
 A/Accession: S02562  
 A/Molecule type: protein  
 A/Residues: 17-166 <HAY>  
 A/Accession: S02563  
 A/Molecule type: protein  
 A/Residues: 47-71 <HAY>  
 R/Schlesinger, D.H.; Hay, D.I.  
 Int. J. Pept. Protein Res. 27, 373-379, 1986  
 A/Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of the proline-rich phosphoprotein, PRP-1  
 A/Reference number: JF0106; MUID:86222916; PMID:3710693  
 A/Accession: JF0106  
 A/Molecule type: protein  
 A/Residues: 17-161, Q', 163-166 <SC3>  
 R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure to the primary structure and possible origin of the non-glycosylated basic proline-rich phosphoprotein  
 A/Reference number: A38355; MUID:91190884; PMID:1849422  
 A/Accession: G38355  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 123-166 <KAU>  
 R/Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A/Title: Primary structure and possible origin of the non-glycosylated basic proline-rich phosphoprotein  
 A/Reference number: S06153; MUID:90088384; PMID:2688632  
 A/Accession: S06153  
 A/Molecule type: protein  
 A/Residues: 123-166 <ROB>  
 C/Comment: The proposed biological functions are a highly potent inhibitor of crystal growth and a highly potent inhibitor of crystal growth  
 C/Genetics:  
 A/Gene: GDB:PRH2  
 A/Cross-references: GDB:119516; OMIM:168790  
 A/Map position: 12p13.2-12p13.2  
 A/Introns: 22/1, 34/1  
 C/Superfamily: proline-rich protein  
 C/Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva  
 F1-16/Domain: signal sequence #status predicted <SIG>  
 F17-166/Product: protein C #status experimental <PRC>  
 F17-122/Product: protein A #status experimental <PRA>  
 F17-46/Region: aptatic mineral binding  
 F17-71/Product: PRP-3 #status experimental <PRP3>  
 F123-166/Product: peptide P-C #status experimental <PPC>  
 F17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
 F124/38/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 100.0%; Score 34; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
 Db 122 RGRPQG 127

RESULT 2  
 B25372

salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human

C/Species: Homo sapiens (man)  
 C/Date: 29-Aug-1987 #sequence revision 29-Aug-1987 #text change 20-Aug-1999  
 C/Accession: B25372; A57868; S02562; G38355; S06153; B27307  
 R/Isomura, S.; Saitoh, E.; Sanada, K.  
 J. Biochem. 87, 1071-1077, 1980  
 A/Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relationship to the proline-rich phosphoprotein precursor PRH1 (allele PIF)  
 A/Reference number: A91954; MUID:80227634; PMID:77390979  
 A/Contents: Peptide P-C  
 A/Accession: A91954  
 A/Molecule type: protein  
 A/Residues: 123-166 <ISE>  
 R/Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A/Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A/Reference number: S02562; MUID:89061650; PMID:3196309  
 A/Accession: S02562  
 A/Molecule type: protein  
 A/Residues: 17-166 <HAY>  
 A/Accession: S02563  
 A/Molecule type: protein  
 A/Residues: 47-71 <HAY>  
 R/Schlesinger, D.H.; Hay, D.I.  
 Int. J. Pept. Protein Res. 27, 373-379, 1986  
 A/Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of the proline-rich phosphoprotein, PRP-1  
 A/Reference number: JF0106; MUID:86222916; PMID:3710693  
 A/Accession: JF0106  
 A/Molecule type: protein  
 A/Residues: 17-161, Q', 163-166 <SC3>  
 R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure to the primary structure and possible origin of the non-glycosylated basic proline-rich phosphoprotein  
 A/Reference number: A38355; MUID:91190884; PMID:1849422  
 A/Accession: G38355  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 123-166 <KAU>  
 R/Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A/Title: Primary structure and possible origin of the non-glycosylated basic proline-rich phosphoprotein  
 A/Reference number: S06153; MUID:90088384; PMID:2688632  
 A/Accession: S06153  
 A/Molecule type: protein  
 A/Residues: 123-166 <ROB>  
 C/Comment: The proposed biological functions are a highly potent inhibitor of crystal growth and a highly potent inhibitor of crystal growth  
 C/Genetics:  
 A/Gene: GDB:PRH2  
 A/Cross-references: GDB:119516; OMIM:168790  
 A/Map position: 12p13.2-12p13.2  
 A/Introns: 22/1, 34/1  
 C/Superfamily: proline-rich protein  
 C/Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva  
 F1-16/Domain: signal sequence #status predicted <SIG>  
 F17-166/Product: protein C #status experimental <PRC>  
 F17-122/Product: protein A #status experimental <PRA>  
 F17-46/Region: aptatic mineral binding  
 F17-71/Product: PRP-3 #status experimental <PRP3>  
 F123-166/Product: peptide P-C #status experimental <PPC>  
 F17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
 F124/38/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 100.0%; Score 34; DB 2; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
 Db 122 RGRPQG 127

RESULT 3  
 A27307

proline-rich phosphoprotein (gene PRH1, Db allele) - human  
 N/Alternate names: salivary acidic proline-rich protein  
 C/Species: Homo sapiens (man)

C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 29-Aug-1997  
 C;Accession: A27307  
 R;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
 A;Reference number: A27307; MUID:88074309; PMID:3687941  
 A;Accession: A27307  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 1-171 <AZE>  
 A;Cross-references: EMBL:X03203  
 C;Genetics:  
 A;Gene: GDB:PRH1  
 A;Cross-references: GDB:I119515; OMIM:168730  
 A;Map position: 12p13.2-12p13.2  
 C;Superfamily: proline-rich protein  
 C;Keywords: phosphoprotein

Query Match 100.0%; Score 34; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
 Db 127 RGRPQG 132

RESULT 4  
 A70984  
 Probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)  
 C;Species: Mycobacterium tuberculosis  
 C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 21-Jul-2000  
 C;Accession: A70984  
 R;Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 Rajandream, M.A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.  
 Nature 393, 537-544, 1998  
 A;Authors: Squires, R.; Sultston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; MUID:98295987; PMID:9634230  
 A;Accession: A70984  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-2188 <COL>  
 A;Cross-references: GB:Z83957; GB:AL123456; NID:G3242252; PIDN:CAR06099.1; PID:G3242253  
 A;Experimental source: strain H37RV  
 C;Genetics:  
 A;Gene: ppsC  
 C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hcnase homology; [acyl-carrier-protein] S-malonyltransferase homology  
 C;Keywords: carrier protein; phosphopantetheine; phosphoprotein  
 F:56-460/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>  
 F:572-853/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>  
 F:1482-1770/Domain: long-chain alcohol dehydrogenase homology <LADH>  
 F:1802-1982/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 F:2072-2141/Domain: acyl carrier protein homology <ACP1>  
 F:2105/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 100.0%; Score 34; DB 2; Length 2188;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
 Db 1479 RGRPQG 1484

RESULT 5  
 NKV98C  
 core protein p19 - artichoke mottled crinkle virus  
 C;Species: artichoke mottled crinkle virus  
 C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 23-Jul-1999  
 C;Accession: S08430

R;Grieco, P.; Gallitelli, D.  
 Nucleic Acids Res. 18, 1300, 1990  
 A;Title: Nucleotide sequence of the 3'-terminal region of artichoke mottled crinkle tc  
 A;Reference number: S08428; MUID:90206906; PMID:2320427  
 A;Accession: S08430  
 A;Status: translation not shown  
 A;Molecule type: genomic RNA  
 A;Residues: 1-172 <GRI>  
 A;Cross-references: EMBL:X51456; NID:G59016; PIDN:CAA35823.1; PID:G59019  
 C;Superfamily: tombusvirus core protein p19  
 C;Keywords: core protein

Query Match 91.2%; Score 31; DB 1; Length 172;  
 Best Local Similarity 83.3%; Pred. No. 45;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
 Db 158 RGRPEG 163

RESULT 6  
 C95286  
 Probable transposase [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid pSy  
 C;Species: Sinorhizobium meliloti  
 C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C;Accession: C95286  
 R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bo  
 Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9893-9898, 2001  
 A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melil  
 A;Reference number: A95262; MUID:21396509; PMID:11481432  
 A;Accession: C95286  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-195 <KUR>  
 A;Cross-references: GB:AE006469; PIDN:AAK64853.1; PID:G14523267; GSPDB:GN00165  
 A;Experimental source: strain 1021, megaplasmid pSYNA  
 R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Huble  
 pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lela  
 heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
 A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A;Reference number: A96039; MUID:21368234; PMID:11474104  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: SMA0360  
 A;Genome: plasmid

Query Match 91.2%; Score 31; DB 2; Length 195;  
 Best Local Similarity 83.3%; Pred. No. 51;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
 Db 41 RGRPEG 46

RESULT 7  
 S76077  
 hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C;Species: Synechocystis sp.  
 A;Variety: PCC 6803  
 C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C;Accession: S76077  
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N  
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu  
 DNA Res. 3, 109-136, 1996  
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocyst  
 s.  
 A;Reference number: S74322; MUID:97061201; PMID:8905231

```

A:Accession: S76077
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-284 <KAN>
A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAAL0055.1; PID:d101070
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match      91.2%; Score 31; DB 2; Length 284;
Best Local Similarity 83.3%; Pred. No. 71;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 RGRPQG 6
Db      246 KGRPQG 251

RESULT 8
S11034
gene P1 protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C:Accession: S11034
R:Maschat, F.; Dubertret, M.L.; Therond, P.; Claverie, J.M.; Lepesant, J.A.
J. Mol. Biol. 214, 359-372, 1990
A:Title: Structure of the ecdysone-inducible P1 gene of Drosophila melanogaster.
A:Reference number: S11034; MUID:90339481; PMID:1696317
A:Accession: S11034
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1030 <NAS>
A:Cross-references: EMBL:X69965; NID:g7960; PIDN:CAA49588.1; PID:g7961
C:Genetics:
A:Gene: FlyBase:Fbp1
A:Cross-references: FlyBase:FBgn0000639
A:Introns: 31/1

Query Match      91.2%; Score 31; DB 2; Length 1030;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 RGRPQG 6
Db      955 KGRPEG 960

RESULT 9
S21010
filamentous hemagglutinin B precursor - Bordetella pertussis
C:Species: Bordetella pertussis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S21010; S18225; S10236; S62090; A41475; S12465
R:Reiman, D.A.
submitted to the EMBL Data Library, March 1990
A:Reference number: S12465
A:Accession: S21010
A:Molecule type: DNA
A:Residues: 1-3591 <REL>
A:Cross-references: EMBL:X52156; NID:g39739; PID:g39740
R:Scarlato, V.; Prugnola, A.; Arico, B.; Rappuoli, R.
Mol. Microbiol. 5, 2493-2498, 1991
A:Title: The bvg-dependent promoters show similar behaviour in different Bordetella species
A:Reference number: S18223; MUID:92167814; PMID:1791761
A:Accession: S18225
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-35 <SCA>
R:Domenighini, M.; Reiman, D.; Capiou, C.; Falkow, S.; Prugnola, A.; Scarlato, V.; Rappuoli, R.
Mol. Microbiol. 4, 787-800, 1990
A:Title: Genetic characterization of Bordetella pertussis filamentous haemagglutinin: a
A:Reference number: S10236; MUID:90355839; PMID:2388559
A:Accession: S10236
A:Molecule type: DNA
A:Residues: 1-3329; AQLPPRPVA, 3340-3591 <DOM>

A:Cross-references: GB:M60351; EMBL:J04531; NID:g144047; PIDN:AAA22974.1; PID:g144048
A:Accession: S62090
A:Molecule type: protein
A:Residues: 1074-1092 <DOS>
R:Delisse-Gathoye, A.M.; Lochet, C.; Jacob, F.; Raaschou-Nielsen, M.; Heron, I.; Ruelle
Infect. Immun. 58, 2895-2905, 1990
A:Title: Cloning, partial sequence, expression, and antigenic analysis of the filament
A:Reference number: A41475; MUID:90354052; PMID:1696934
A:Accession: A41475
A:Molecule type: DNA
A:Residues: 1-279; R', 281-283; A', 285-515; V', 517-613; G', 615-1087 <DEL>
A:Cross-references: GB:X53405; NID:g39737; PIDN:CAA37481.1; PID:g39738
C:Genetics:
A:Gene: fhaB; fha
C:Superfamily: Bordetella filamentous hemagglutinin B
C:Keywords: hemagglutinin

Query Match      91.2%; Score 31; DB 1; Length 3591;
Best Local Similarity 83.3%; Pred. No. 7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 RGRPQG 6
Db      2189 RGRPEG 2194

RESULT 10
JC5931
high mobility group I HMGI chromosomal protein isoform C-alpha - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C:Accession: JC5931
R:Kottickal, L.V.; Sarada, B.; Ashar, H.; Chada, K.; Nagarajan, L.
Biochem. Biophys. Res. Commun. 242, 452-456, 1998
A:Title: Preferential expression of HMGI-C isoforms lacking the acidic carboxy termina
A:Reference number: JC5931; MUID:98113374; PMID:9446816
A:Accession: JC5931
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-90 <KOT>
C:Superfamily: nonhistone chromosomal protein HMGI-I

Query Match      88.2%; Score 30; DB 2; Length 90;
Best Local Similarity 83.3%; Pred. No. 40;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 RGRPQG 6
Db      49 RGRPKG 54

RESULT 11
A31895
nonhistone chromosomal protein HMGI-I(Y) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 23-Jul-1999
C:Accession: A31895
R:Johnson, K.R.; Lehn, D.A.; Elton, T.S.; Barr, P.J.; Reeves, R.
J. Biol. Chem. 263, 18338-18342, 1988
A:Title: Complete murine cDNA sequence, genomic structure, and tissue expression of th
A:Reference number: A31895; MUID:89054013; PMID:3192537
A:Accession: A31895
A:Molecule type: mRNA
A:Residues: 1-96 <JOH>
A:Cross-references: GB:J04179; NID:g193883; PIDN:AAA37820.1; PID:g309298
C:Superfamily: nonhistone chromosomal protein HMGI-I
C:Keywords: alternative splicing; chromosomal protein; DNA binding; nucleus; phosphor

Query Match      88.2%; Score 30; DB 2; Length 96;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 RGRPQG 6

```

```

Db          47 RGRPKG 52
|||||:|
RESULT 12
JC5933
high mobility group I HMGI chromosomal protein isoform C-gamma - human
C:Species: Homo sapiens (man)
C>Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C:Accession: JC5933
R:Kottickal, L.V.; Sarada, B.; Ashar, H.; Chada, K.; Nagarajan, L.
Biochem. Biophys. Res. Commun. 242, 452-456, 1998
A:Title: Preferential expression of HMGI-C isoforms lacking the acidic carboxy terminal
A:Reference number: JC5931; MUID:98113374; PMID:9446816
A:Accession: JC5933
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-96 <KOT>
C:Superfamily: nonhistone chromosomal protein HMGI-I
Query Match      88.2%; Score 30; DB 2; Length 96;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPKG 6
|||||:|
Db          49 RGRPKG 54
|||||:|
RESULT 13
JC5932
high mobility group I HMGI chromosomal protein isoform C-beta - human
C:Species: Homo sapiens (man)
C>Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C:Accession: JC5932
R:Kottickal, L.V.; Sarada, B.; Ashar, H.; Chada, K.; Nagarajan, L.
Biochem. Biophys. Res. Commun. 242, 452-456, 1998
A:Title: Preferential expression of HMGI-C isoforms lacking the acidic carboxy terminal
A:Accession: JC5932
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-105 <KOT>
C:Superfamily: nonhistone chromosomal protein HMGI-I
Query Match      88.2%; Score 30; DB 2; Length 105;
Best Local Similarity 83.3%; Pred. No. 46;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPKG 6
|||||:|
Db          49 RGRPKG 54
|||||:|
RESULT 14
A32794
nonhistone chromosomal protein HMGI-I - human
N:Alternate names: high-mobility group protein HMGI-I; HMGIY splice form I
N:Contains: nonhistone chromosomal protein HMGI-Y
C:Species: Homo sapiens (man)
C>Date: 08-Dec-1989 #sequence_revision 08-Dec-1989 #text_change 23-Jul-1999
C:Accession: A32794; B32794; S05322; S41769; A26907; A32250
R:Johnson, K.R.; Lehn, D.A.; Reeves, R.
Mol. Cell. Biol. 9, 2114-2123, 1989.
A:Title: Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-gr
A:Reference number: A32794; MUID:89313758; PMID:2701943
A:Accession: A32794
A:Molecule type: mRNA
A:Residues: 1-107 <JOH>
A:Cross-references: GB:M23614; NID:g184246; PIDN:AAA88072.1; PID:g306866; GB:J04239
A:Accession: B32794
A:Molecule type: mRNA
A:Residues: 1-34,46-107 <J02>

A:Cross-references: GB:M23615; NID:g184252; PIDN:AAA88073.1; PID:g306870; GB:J04239
R:Eckner, R.; Birnstiel, M.L.
Nucleic Acids Res. 17, 5947-5959, 1989
A:Title: Cloning of cDNAs coding for human HMGI and HMGI Y proteins: both are capable
A:Reference number: S05321; MUID:89366631; PMID:2505228
A:Accession: S05322
A:Molecule type: mRNA
A:Residues: 1-107 <ECK>
A:Cross-references: EMBL:X14957; NID:g32336; PIDN:CAA33080.1; PID:g32337
R:Friedman, M.; Holth, L.T.; Zoghbi, H.Y.; Reeves, R.
Nucleic Acids Res. 21, 4259-4267, 1993
A:Title: Organization, inducible-expression and chromosome localization of the human I
A:Reference number: S41769; MUID:94021372; PMID:8414980
A:Accession: S41769
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-107 <FRI>
A:Cross-references: EMBL:L17131; NID:g306868; PIDN:AA00145.1; PID:g306869
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1993
R:Lund, T.; Dahl, K.H.; Mork, E.; Holtlund, J.; Laland, S.G.
Biochem. Biophys. Res. Commun. 158, 646-651, 1989
A:Title: The amino acid sequence of the chromosomal protein HMGI-Y, its relation to HMGI
A:Reference number: A32250; MUID:89149775; PMID:2320035
A:Accession: A32250
A:Molecule type: protein
A:Residues: 4-25, 'X', 27-68, 69-107, 'Q' <LUN>
R:Karlsson, J.R.; Mork, E.; Holtlund, J.; Laland, S.G.; Lund, T.
Biochem. Biophys. Res. Commun. 158, 646-651, 1989
A:Title: The amino acid sequence of the chromosomal protein HMGI-Y, its relation to HMGI
A:Reference number: A32250; MUID:89149775; PMID:2320035
A:Accession: A32250
A:Molecule type: protein
A:Residues: 4-34, 46-107, 'QQ' <KAR>
C:Genetics:
A:Gene: GDB:HMGIY
A:Cross-references: GDB:134205; OMIM:600701
A:Map position: 6p21-6p21
A:Introns: 45/3; 73/3; 90/3
C:Superfamily: nonhistone chromosomal protein HMGI-I
C:Keywords: alternative splicing; blocked amino end; chromosomal protein; DNA binding;
F:1-107/Product: nonhistone chromosomal protein HMGI-I #status predicted <MATI>
F:1-34, 46-107/Product: nonhistone chromosomal protein HMGI-Y #status predicted <MATY>
F:102,103/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match      88.2%; Score 30; DB 2; Length 107;
Best Local Similarity 83.3%; Pred. No. 47;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPKG 6
|||||:|
Db          58 RGRPKG 63
|||||:|
RESULT 15
JC4575
high-mobility-group I protein C - mouse
N:Alternate names: nuclear phosphoprotein
C:Species: Mus musculus (house mouse)
C>Date: 10-Apr-1996 #sequence_revision 24-May-1996 #text_change 23-Jul-1999
C:Accession: JC4575; S22597
R:Manfioletti, G.; Rustighi, A.; Mantovani, F.; Goodwin, G.H.; Giaccotti, V.
Gene 167, 249-253, 1995
A:Title: Isolation and characterization of the gene coding for murine high-mobility-g
A:Reference number: JC4575; MUID:96144283; PMID:8566786
A:Accession: JC4575
A:Molecule type: mRNA
A:Residues: 1-108 <MAN>
A:Cross-references: GB:L41617
R:Manfioletti, G.; Giaccotti, V.; Bandiera, A.; Buratti, E.; Sautiere, P.; Cary, P.; C
Nucleic Acids Res. 19, 6793-6797, 1991
A:Title: cDNA cloning of the HMGI-C phosphoprotein, a nuclear protein associated with
A:Reference number: S22597; MUID:92107664; PMID:1762909
A:Accession: S22597

```

A:Molecule type: mRNA  
A:Residues: 1-108 <MA2>  
A:Cross-references: EMBL:X58380; NID:G51343; PIDN:CAA41270.1; PID:G51344  
C:Comment: This protein is a small, highly charged protein, and contains an important co-  
factor in combinatorial promoters, and involved in the process of cell transformation. It  
C:Genetics:  
A:Gene: Hmgj-C  
A:Introns: 37/3; 56/3; 83/3; 94/3  
C:Superfamily: nonhistone chromosomal protein HMG-I  
C:Keywords: DNA binding; phosphoprotein  
F:1-37,38-66,67-83,84-94/Domain: DNA binding #status predicted <DNB>  
F:84-94/Region: acidic  
  
Query Match 88.2%; Score 30; DB 2; Length 108;  
Best Local Similarity 83.3%; Pred.No: 47;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 RGRPQG 6  
|||  
Db 49 RGRPKG 54  
  
Search-completed: April 6, 2004, 16:16:50  
Job time : 7.72897 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 3.86916 Seconds  
(without alignments)

80.746 Million cell updates/sec

Title: US-10-009-709-2

Perfect score: 34

Sequence: 1 RGRPQG 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	166	1 PRPC_HUMAN	P02810 homo sapien
2	31	91.2	172	1 VP19_AMCV	P15960 artichoke m
3	31	91.2	172	1 VP19_TBVA	P50625 tomato bush
4	31	91.2	1030	1 FBPI_DROME	O04891 drosophila
5	31	91.2	3590	1 FBAB_BOPE	P12855 bordetella
6	30	88.2	95	1 HWGY_MOUSE	P17095 mus musculus
7	30	88.2	106	1 HMGI_CRIGR	Q9qxp3 cricetus
8	30	88.2	106	1 HMGI_HUMAN	P17096 homo sapien
9	30	88.2	108	1 HMGC_MOUSE	P52927 mus musculus
10	30	88.2	109	1 HMGC_HUMAN	P52926 homo sapien
11	30	88.2	218	1 RRAS_HUMAN	P10301 homo sapien
12	30	88.2	218	1 RRAS_MOUSE	P10833 mus musculus
13	30	88.2	261	1 PRP2_MOUSE	P05142 mus musculus
14	30	88.2	264	1 RS31_ARATH	P92964 arabidopsis
15	30	88.2	484	1 MEC2_MOUSE	Q922d6 mus musculus
16	30	88.2	486	1 MEC2_HUMAN	P15608 homo sapien
17	30	88.2	492	1 MEC2_RAT	Q00566 rattus norv
18	30	88.2	510	1 YK67_CASEL	P34528 caenorhabdi
19	30	88.2	535	1 V050_HSVSA	Q01012 herpesvirus
20	30	88.2	693	1 SUV7_ARATH	Q9csp1 arabidopsis
21	30	88.2	740	1 CEB1_HUMAN	Q15027 homo sapien
22	30	88.2	740	1 CAG_SMRVH	P21411 squirrel mo
23	30	88.2	755	1 SUV8_ARATH	Q9csp0 arabidopsis
24	30	88.2	901	1 CPBA_ANASP	P58572 anabaena sp
25	30	88.2	901	1 CPBA_ANAVA	O86109 anabaena va
26	30	88.2	1069	1 C110_HUMAN	Q9nb22 homo sapien
27	30	88.2	2164	1 CCAA_MOUSE	P97445 mus musculus
28	30	88.2	2212	1 CCAA_RAT	P54282 rattus norv
29	30	88.2	2424	1 CCAA_RABIT	P27884 coryctolagus
30	30	88.2	2505	1 CCAA_HUMAN	O00555 homo sapien
31	29	85.3	53	1 NCE1_YEAST	Q02820 saccharomyc
32	29	85.3	59	1 PART_HUMAN	Q9npd0 homo sapien
33	29	85.3	83	1 EX7S_RHIME	Q92r19 rhizobium m

RESULT 1					
PRPC_HUMAN					
ID	PRPC_HUMAN	STANDARD;	PRT;	166 AA.	
AC	P02810;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	01-MAR-1989 (Rel. 10, Last sequence update)				
DT	15-MAR-2004 (Rel. 43, Last annotation update)				
DE	Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (protein A/protein C) [contains: Peptide P-C].				
DE	Peptide P-C.				
GN	PRH1 AND PRH2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).				
RX	MEDLINE=86196106; PubMed=3009472;				
RA	Kim H.-S., Maeda N.;				
RT	"Structures of two Haell-type genes in the human salivary proline-rich protein multigene family.";				
RL	J. Biol. Chem. 261:6712-6718(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).				
RX	MEDLINE=85289325; PubMed=2993301;				
RA	Maeda N., Kim H.-S., Azen E.A., Smithies O.;				
RT	"Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.";				
RL	J. Biol. Chem. 260:11123-11130(1985).				
RN	[3]				
RP	SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).				
RX	MEDLINE=89061650; PubMed=3196309;				
RA	Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K., Madapallimattam G., Schluckebier S.K.;				
RT	"The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";				
RL	Biochem. J. 255:15-21(1988).				
RN	[4]				
RP	SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).				
RX	MEDLINE=88074309; PubMed=3687941;				
RA	Azen E.A., Kim H.-S., Goodman P., Flynn S., Maeda N.;				
RT	"Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich proteins Pa, Db, and PIF.";				
RL	Am. J. Hum. Genet. 41:1035-1047(1987).				
RN	[5]				
RP	SEQUENCE OF 17-166 (PRP-2).				
RX	MEDLINE=86222916; PubMed=3710693;				
RA	Schlesinger D.H., Hay D.I.;				
RT	"Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva.";				
RL	Int. J. Pept. Protein Res. 27:373-379(1986).				
RN	[6]				
RP	SEQUENCE OF 17-166 (PROTEIN C).				
RX	MEDLINE=80204368; PubMed=7380845;				
RA	Wong R.S.C., Bennick A.;				

P81489 homo sapien  
Q88ck5 pseudomonas  
Q889ul pseudomonas  
P88428 saccharomyc  
Q9gy17 mus musculus  
P05143 mus musculus  
P41155 mus musculus  
Q9jml0 marmota mon  
Q53956 streptomyces  
P39567 saccharomyc  
P36049 saccharomyc  
Q9a2h2 caulobacter

RT "The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A.";  
RL J. Biol. Chem. 255:5943-5948 (1980).  
[7]  
RN  
RP SEQUENCE OF 17-46 (PROTEIN C).  
RX MEDLINE=81191179; PubMed=7228490;  
RA Schlesinger D.H., Hay D.I.;  
RT "Primary structure of the active tryptic fragments of human and monkey salivary anionic proline-rich proteins.";  
RL Int. J. Pept. Protein Res. 17:34-41 (1981).  
[8]  
RN  
RP SEQUENCE OF 17-122 (PROTEIN A).  
RX MEDLINE=79173237; PubMed=438215;  
RA Wong R.S.C., Hofmann I., Bennick A.;  
RT "The complete primary structure of a proline-rich phosphoprotein from human saliva.";  
RL J. Biol. Chem. 254:4800-4808 (1979).  
[9]  
RN  
RP SEQUENCE OF 17-122 (PROTEIN A).  
RA Schlesinger D.H., Hay D.I.;  
RT "Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent inhibitor of calcium phosphate precipitation in human parotid saliva.";  
RL (In) Gross E., Meienhofer J. (eds.);  
RL Peptides: structure and biological function (Proceedings of the 6th American peptide symposium), pp.133-136, Pierce Chemical Co., Rockford IL. (1979).  
[10]  
RN  
RP SEQUENCE OF 123-166 (PEPTIDE P-C).  
RX MEDLINE=80227634; PubMed=7390979;  
RA Isemura S., Saitoh E., Sanada K.;  
RT "The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to a salivary proline-rich phosphoprotein, protein C.";  
RL J. Biochem. 87:1071-1077 (1980).  
[11]  
RN  
RP VARIANT PRH2-3 LYS-163.  
RA Azen E.A.;  
RT "A frequent mutation in the acidic proline-rich protein gene, PRH2, causing a Q147K change closely adjacent to the bacterial binding domain of the cognate salivary PRP (Pr1') in Afro-Americans.";  
RL Hum. Mutat. 12:72-72 (1998).  
CC -!- FUNCTION: PRP's act as highly potent inhibitors of crystal growth of calcium phosphates. They provide a protective and reparative environment for dental enamel which is important for the integrity of the teeth.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIP-S yield PRP-4, PRP-3 (protein A), and PIP-F, respectively.  
CC -!- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele; allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is also known as PR1'.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; K03202; AAA60183.1; -;  
CC EMBL; K03203; AAA60184.1; -;  
CC EMBL; M13057; AAA98807.1; -;  
CC EMBL; M13058; AAA98808.1; -;  
CC Genew; HGNC:9366; PRH1.  
CC Genew; HGNC:9367; PRH2.  
CC MIM; 168730; -;  
CC MIM; 168730; -;  
CC MIM; 168710; -;  
CC GO; GO:0005615; Extracellular space; TAS.  
CC Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism;  
KW Pyrrolidone carboxylic acid.

FT SIGNAL 1 16  
FT CHAIN 17 166  
FT CHAIN 17 122  
FT CHAIN 17 46  
FT CHAIN 123 166  
FT DOMAIN 17 46  
FT MOD\_RES 17 17  
FT MOD\_RES 24 24  
FT MOD\_RES 38 38  
FT VARIANT 20 20  
FT VARIANT 66 66  
FT VARIANT 163 163  
FT CONFLICT 41 41  
FT SEQUENCE 166 AA; 17017 MW; A7DF62BF94E3C3EF CRC64;  
Query Match 100.0%; Score 34; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 4.1; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;  
QY 1 RGRPQG 6  
DB 122 RGRPQG 127  
RESULT 2  
VF19\_AMCV STANDARD; PRT; 172 AA.  
AC P15960;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Core protein P19.  
OS Artichoke mottled crinkle virus (AMCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
OX Tombusvirus.  
NCBI\_TaxID=12142;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Barri-Dr. Gallitelli isolate;  
MEDLINE=9026806; PubMed=2320427;  
RA Grieco F., Gallitelli D.;  
RT "Nucleotide sequence of the 3'-terminal region of artichoke mottled crinkle tombusvirus RNA.";  
RL Nucleic Acids Res. 18:1300-1300 (1990).  
CC -!- SIMILARITY: Belongs to the tombusviridae P19 family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X51456; CAA35823.1; -;  
CC PIR; S08430; NKVGBG.  
CC InterPro; IPR004905; Tombus\_P19.  
CC Pfam; PF03220; Tombus\_P19; 1.  
CC ProDom; PD005743; Tombus\_P19; 1.  
KW Core protein.  
SQ SEQUENCE 172 AA; 19280 MW; 5EE2911E6DA96901 CRC64;  
Query Match 91.3%; Score 31; DB 1; Length 172;  
Best Local Similarity 83.3%; Pred. No. 18;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRPQG 6  
DB 122 RGRPQG 127

```

DB      158 RGRPEG 163

RESULT 3
VP19_TBSVA STANDARD; PRT; 172 AA.
AC P50625;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Core protein P19.
OS Tomato bushy stunt virus (strain A23) (TBSV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC Tombusvirus.
OX NCBI_TaxID=70156;
RN [1]
RP SEQUENCE FROM N.A.
RA Luis-Arteaga M., Rodriguez-Cerezo E., Fraile A., Saez E.,
RA Garcia-Arenal F.;
RT "Different tomato bushy stunt virus strains cause disease outbreaks on
RT solanaceous crops in Spain."
RL Phytopathology 86:535-542(1996).
CC -1- SIMILARITY: Belongs to the tombusviridae P19 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z68896; CAA93126.1; -
DR InterPro; IPR004905; Tombus_P19.
DR Pfam; PF03220; Tombus_P19; 1.
DR ProDom; PD005743; Tombus_P19; 1.
KW Core protein.
SQ SEQUENCE 172 AA; 19469 MW; 92A2DE57758627FC CRC64;

Query Match 91.2%; Score 31; DB 1; Length 172;
Best Local Similarity 83.3%; Pred. NO. 18;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPEG 6
DB 158 RGRPEG 163

RESULT 4
FBP1_DROME STANDARD; PRT; 1030 AA.
AC Q04691;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Fat-body protein-1 precursor (P1 protein).
DE FBP1 OR P1.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Canton-S; TISSUE=Fat body;
RX MEDLINE=90355839; PubMed=2388559;
RA Maschat F., Dubertret M.-L., Therond P., Claverie J.-M.,
RA Lepesant J.-A.;
RT "Structure of the ecdysone-inducible P1 gene of Drosophila
RT melanogaster."
RL J. Mol. Biol. 214:359-372(1990).
RN [2]
RP SEQUENCE FROM N.A.

DB      158 RGRPEG 163

Query Match 91.2%; Score 31; DB 1; Length 1030;
Best Local Similarity 83.3%; Pred. NO. 11e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPEG 6
DB 955 RGRPEG 960

RESULT 5
PHAB_BORPE STANDARD; PRT; 3590 AA.
AC P12235;
DT 01-OCT-1989 (Rel. 12, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Filamentous hemagglutinin.
GN PHAB OR BP1879.
OS Bordetella pertussis
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90355839; PubMed=2388559;
RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic Characterization of Bordetella pertussis filamentous
RT hemagglutinin: a protein processed from an unusually large
RT precursor."
RL Mol. Microbiol. 4:787-800(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia I. / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,

```

RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
 RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Feilwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
 RA Leather S., Moulé S., Norbertzak H., O'Neill S., Ormond D., Price C.,  
 RA Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,  
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares K.,  
 RA Unwin L., Whitehead S., Bartell B.G., Maskell D.J.  
 RT "Comparative analysis of the genome sequences of *Bordetella pertussis*,  
 RT *Bordetella parapertussis* and *Bordetella bronchiseptica*.";  
 RL Nat. Genet. 35:32-40(2003).  
 RN [3]  
 RP MEDLINE=99202384; PubMed=2539596;  
 RX Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
 RA "Filamentous hemagglutinin of *Bordetella pertussis*: nucleotide  
 RT sequence and crucial role in adherence.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).  
 CC -!- FUNCTION: Evidence for a role in host-cell binding and infection.  
 CC -!- SUBCELLULAR LOCATION: Surface.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M60351; AAA22974.1; --  
 CC EMBL; BX640416; CA542162.1; --  
 CC PIR; S21010; S21010.  
 CC InterPro; IPR008619; Fil haemagg.  
 CC InterPro; IPR008638; Haemagg act.  
 CC Pfam; PF05594; Fil haemagg; 18.  
 CC Pfam; PF05860; Haemagg; act; 1.  
 CC Antigen; Hemagglutinin; Complete proteome.  
 CC CONFLICT 507 508 KQ -> NE (IN REF. 1 AND 3).  
 CC CONFLICT 1454 1454 A -> P (IN REF. 1 AND 3).  
 CC CONFLICT 3574 3590 VEDIGGNVRFVETNK -> SRISAARTTGSSMKPTNR  
 CC (IN REF. 1).  
 CC SEQUENCE 3590 AA; 367519 MW; C00BD8E22C9DB41D CRC64;  
 Query Match 91.28; Score 31; DB 1; Length 3590;  
 Best Local Similarity 83.34; Pred. No. 4.1e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPQG 6  
 DB 2189 RGRPEG 2194  
 RESULT 6  
 HMGY MOUSE  
 ID HMGY MOUSE STANDARD; PRT; 95 AA.  
 AC P17035.  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE High mobility group protein HMGY-y (High mobility group AT-hook 1).  
 GN HMGA1 OR HMGIY OR HMGI.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=89054013; PubMed=3192537;  
 RA Johnson K.R., Lehn D.A., Elton T.S., Barr P.J., Reeves R.;  
 RT "Complete murine cDNA sequence, genomic structure, and tissue  
 RT expression of the high mobility group protein HMGY-1(Y).";  
 RL J. Biol. Chem. 263:18338-18342(1988).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP PHOSPHORYLATION, AND ACETYLATION.  
 RX MEDLINE=93054543; PubMed=1429598;  
 RA Ferranti P., Malorni A., Marino G., Pucci P., Goodwin G.H.,  
 RA Manfioletti G., Giancotti V.;  
 RT "Mass spectrometric analysis of the HMGY protein from Lewis lung  
 RL carcinoma. Identification of phosphorylation sites.";  
 CC J. Biol. Chem. 267:22486-22489(1992).  
 CC -!- FUNCTION: HMGY-I/y bind preferentially to the minor groove of A+T  
 CC rich regions in double stranded DNA. It is suggested that these  
 CC proteins could function in nucleosome phasing and in the 3' end  
 CC processing of mRNA transcripts. They are also involved in the  
 CC transcription regulation of genes containing, or in close  
 CC proximity to A+T-rich regions.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=HMGY-Y;  
 CC IsoId=PI7095-1; Sequence=Displayed;  
 CC Name=HMGI-I;  
 CC IsoId=PI7095-2; Sequence=Not described;  
 CC -!- SIMILARITY: Contains 3 A.T hook DNA-binding repeats.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; J04179; AAA37820.1; --  
 CC EMBL; BC008125; AAH08125.1; --  
 CC PIR; A31895; A31895.  
 CC TRANSPAC; T01852;  
 CC MGD; MGI:96160; Hmgal.  
 CC InterPro; IPR006637; AT hook.  
 CC InterPro; IPR000116; Highmobility\_IV.  
 CC Pfam; PF02178; AT hook; 3.  
 CC PRINTS; PR00929; ATHOOK.  
 CC ProDom; PD005593; Highmobility\_IV; 1.  
 CC SMART; SM00384; AT hook; 3.  
 CC PROSITE; PS00354; HMGI\_Y; 3.  
 CC Acetylation.  
 CC Nuclear protein; Chromosomal protein; DNA-binding; Repeat;  
 CC Transcription regulation; Alternative splicing; Phosphorylation;  
 CC Acetylation.  
 CC INIT MET 0 0  
 CC DNA\_BIND 20 30 A.T HOOK 1.  
 CC DNA\_BIND 41 51 A.T HOOK 2.

```

FT DNA_BIND 66 77 A.T HOOK 3.
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 87 87 PHOSPHORYLATION.
FT MOD_RES 90 90 PHOSPHORYLATION.
FT MOD_RES 91 91 PHOSPHORYLATION.
SQ SEQUENCE 95 AA; 10486 MW; 6E8AA6DESA95F8A5 CRC64;
Query Match 88.2%; Score 30; DB 1; Length 95;
Best Local Similarity 83.3%; Pred. No. 16;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRPOG 6
Db 46 RGRPKG 51

RESULT 7
HMGI_CRIGR STANDARD; PRT; 106 AA.
AC Q9QXP3; Q9QXP2;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE High mobility group protein HMG-I/HMG-Y (HMG-I(Y)) (High mobility
group AT-hook 1).
GN HMG1 OR HMG1Y.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS HMG-I AND HMG-Y).
RA Aldrich T.L., Reeves R., Lee C.C., Thomas J.N., Morris A.E.;
"High mobility group proteins implicated in amplification of CHO cell DNA.";
Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HMG-I/Y bind preferentially to the minor groove of A+T
rich regions in double stranded DNA. It is suggested that these
proteins could function in nucleosome phasing and in the 3' end
processing of mRNA transcripts. They are also involved in the
transcription regulation of genes containing, or in close
proximity to A+T-rich regions (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS.
CC Event-Alternative splicing; Named isoforms=2;
CC Name=HMG-I;
CC IsoId=Q9QXP3-1; Sequence=Displayed;
CC Name=HMG-Y;
CC IsoId=Q9QXP3-2; Sequence=VSP 002181;
CC -!- SIMILARITY: Contains 3 A.T hook DNA-binding repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF193762; AAF06666.2; -
CC EMBL; AF193763; AAF06667.2; -
CC InterPro; IPR000637; AT hook.
CC InterPro; IPR000116; Highmobility_IY.
CC Pfam; PF02178; AT hook; 3.
CC PRINTS; PR00929; ATHOOK.
CC ProDom; PD005593; Highmobility_IY; 1.
CC SMART; SM00384; AT hook; 3.
CC PROSITE; PS00354; HMG1_Y; 3.
CC Nuclear protein; Chromosomal protein; DNA-binding; Repeat;
KW Transcription regulation; Alternative splicing; Phosphorylation;
KW Acetylation.
FT INIT MET 0 0 BY SIMILARITY.
FT DNA_BIND 20 30 A.T HOOK 1.

```

```

FT DNA_BIND 52 62 A.T HOOK 2.
FT DNA_BIND 77 88 A.T HOOK 3.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT MOD_RES 98 98 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 101 101 PHOSPHORYLATION (BY CK) (BY SIMILARITY).
FT MOD_RES 102 102 PHOSPHORYLATION (BY CK) (BY SIMILARITY).
FT VARSPPLIC 34 44 Missing (in isoform HMG-Y).
FT FTId=VSP 002181.
SQ SEQUENCE 106 AA; 11485 MW; 30E05217D3D2790E CRC64;
Query Match 88.2%; Score 30; DB 1; Length 106;
Best Local Similarity 83.3%; Pred. No. 17;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRPOG 6
Db 57 RGRPKG 62

RESULT 8
HMGI_HUMAN STANDARD; PRT; 106 AA.
AC P17036; P10910;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE High mobility group protein HMG-I/HMG-Y (HMG-I(Y)) (High mobility
group AT-hook 1).
GN HMG1 OR HMG1Y.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (HMG-I AND HMG-Y).
RX MEDLINE=89366631; PubMed=2505228;
RA Eckner R., Birnstiel M.L.;
RT "Cloning of cDNAs coding for human HMG I and HMG Y proteins: both are
capable of binding to the octamer sequence motif.";
RL Nucleic Acids Res. 17:5947-5959(1989).
RN [2]
SEQUENCE FROM N.A. (HMG-I AND HMG-Y).
RX MEDLINE=89313759; PubMed=2701943;
RA Johnson K.R., Lehn D.A., Reeves R.;
RT "Alternative processing of mRNAs encoding mammalian chromosomal high-
mobility-group proteins HMG-I and HMG-Y.";
RL Mol. Cell. Biol. 9:2114-2123(1989).
RN [3]
SEQUENCE FROM N.A. (HMG-I AND HMG-Y).
RP TISSUE=Placenta;
RX MEDLINE=94021372; PubMed=8414980;
RA Friedmann M., Holth L.T., Zoghbi H.Y., Reeves R.;
RT "Organization, inducible-expression and chromosome localization of
the human HMG-I(Y) nonhistone protein gene.";
RL Nucleic Acids Res. 21:4259-4267(1993).
RN [4]
SEQUENCE FROM N.A. (ISOFORMS HMG-I AND HMG-Y).
RP TISSUE=B-cell, and Uterus;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fanev J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

```

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Kzywinski M.I., Skalska U., Smalilus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP SEQUENCE OF 3-106 (HMG-Y).  
 RX MEDLINE=87298501; PubMed=3619901;  
 RA Lund T., Dahl K.H., Mork E., Holtlund J., Laland S.G.;  
 RA "The human chondrosomal protein HMG I contains two identical  
 RT palindromic amino acid sequences";  
 RL Biochem. Biophys. Res. Commun. 146:725-730(1987).  
 RN [6]  
 RP SEQUENCE OF 3-106 (HMG-Y).  
 RX MEDLINE=89149775; PubMed=2920035;  
 RA Karlson J.R., Mork E., Holtlund J., Laland S.G., Lund T.;  
 RA "The amino acid sequence of the chondrosomal protein HMG-Y, its  
 RT relation to HMG-I and possible domains for the preferential binding  
 RT of the proteins to stretches of A-T base pairs";  
 RL Biochem. Biophys. Res. Commun. 158:646-651(1989).  
 RN [7]  
 RP DNA-BINDING DOMAINS  
 RX MEDLINE=90256776; PubMed=1692833;  
 RA Reeves R., Nissen M.S.;  
 RA "The A.T-DNA-binding domain of mammalian high mobility group I  
 RT chondrosomal proteins. A novel peptide motif for recognizing DNA  
 RT structure";  
 RL J. Biol. Chem. 265:8573-8582(1990).  
 RN [8]  
 RP CHONDROSOMAL TRANSLOCATION WITH RAD51L1.  
 RX MEDLINE=21974627; PubMed=11978964;  
 RA Blank C., Schoenmakers E.F., Rogalla P., Huys E.H., van Rijk A.A.,  
 RA Drieschner N., Bullerdiek J.;  
 RA "Intragenic breakpoint within RAD51L1 in a t(6;14) (p21.3;q24) of a  
 RT pulmonary chondroid hamartoma";  
 RL Cytogenet. Cell Genet. 95:17-19(2001).  
 RN [9]  
 RP STRUCTURE BY NMR OF 50-74 AND 79-88.  
 RX MEDLINE=97397355; PubMed=9253416;  
 RA Ruth J.R., Bewley C.A., Nissen M.S., Evans J.N., Reeves R.,  
 RA Gronenborn A.M., Clore G.M.;  
 RA "The solution structure of an HMG-I(Y)-DNA complex defines a new  
 RT architectural minor groove binding motif";  
 RL Nat. Struct. Biol. 4:657-665(1997).  
 CC -!- FUNCTION: HMG-I/Y bind preferentially to the minor groove of A-T  
 CC rich regions in double stranded DNA. It is suggested that these  
 CC proteins could function in nucleosome phasing and in the 3' end  
 CC processing of mRNA transcripts. They are also involved in the  
 CC transcription regulation of genes containing, or in close  
 CC proximity to A-T-rich regions.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=HMG-I;  
 CC IsoId=p17096-1; Sequence=Displayed;  
 CC Name=HMG-Y;  
 CC IsoId=p17096-2; Sequence=VSP\_002182;  
 CC -!- DISEASE: Involved in pulmonary chondroid hamartoma through a  
 CC chromosomal translocation t(6;14)(p21;q23-24) that involves HMGAI  
 CC and RAD51L1.  
 CC -!- SIMILARITY: Contains 3 A-T hook DNA-binding repeats.  
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;  
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/HMGYID221.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X14957; CAA33080.1; -  
 DR EMBL; X14956; CAA33081.1; -  
 DR EMBL; M23614; AAA88072.1; -  
 DR EMBL; M23615; AAA88073.1; -  
 DR EMBL; M23616; AAA88074.1; -  
 DR EMBL; M23617; AAA88075.1; -  
 DR EMBL; M23618; AAA88076.1; -  
 DR EMBL; M23619; AAA35998.1; -  
 DR EMBL; L17131; AAB00145.1; -  
 DR EMBL; BC004924; AAH04924.1; -  
 DR EMBL; BC008832; AAH08832.1; -  
 DR PIR; A32794; A32794.  
 DR PDB; 2EZD; 15-OCT-97.  
 DR PDB; 2EZB; 15-OCT-97.  
 DR PDB; 2EZG; 15-OCT-97.  
 DR TRANSFAC; T01851; -  
 DR TRANSFAC; T01980; -  
 DR MIM; 600701; -  
 DR GO; GO:0003680; F:AT DNA binding; TAS.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.  
 DR InterPro; IPR00637; AT hook.  
 DR Pfam; PF02178; AT hook; 3.  
 DR PRINTS; PR00929; ATHOOK.  
 DR PRODOM; PD005593; Highmoblty\_IY; 1.  
 DR SMART; SMC0384; AT hook; 3.  
 DR PROSITE; PS00354; HMG1\_Y; 3.  
 KW Nuclear protein; Chondrosomal  
 KW transcription regulation; Alternative splicing; Phosphorylation;  
 KW Acetylation; Chromosomal translocation; 3D-structure.  
 FT INIT MET 0 0  
 FT DNA\_BIND 20 30 A.T HOOK 1.  
 FT DNA\_BIND 52 62 A.T HOOK 2.  
 FT DNA\_BIND 77 88 A.T HOOK 3.  
 FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 FT MOD\_RES 98 98 PHOSPHORYLATION.  
 FT MOD\_RES 101 101 PHOSPHORYLATION (BY CK).  
 FT MOD\_RES 102 102 PHOSPHORYLATION (BY CK).  
 FT VARSPLIC 34 44 Missing (in isoform HMG-Y).  
 FT CONFLICT 34 34 /FTID=VSP\_002182.  
 FT CONFLICT 106 106 MISSING (IN CLONE 11D).  
 FT CONFLICT 106 106 Q -> QQQ (IN REF. 6).  
 SQ SEQUENCE 106 AA; 11545 MW; 2D568967D3C6284B CRC64;  
 Query Match 88.2%; Score 30; DB 1; Length 106;  
 Best Local Similarity 83.3%; Pred. No. 17;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RGRPOG 6  
 Db 57 RGRPKG 62  
 RESULT 9  
 HMG1\_MOUSE  
 ID\_HMG1\_MOUSE STANDARD; PRT; 108 AA.  
 AC PS2927;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE High mobility group protein HMG1-C (High mobility group AT-hook 2).  
 GN HMG2 OR HMGIC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92107664; PubMed=1762909;  
 RA Manfioletti G., Giannotti V., Bandiera A., Buratti E., Sautiere P.,  
 RA Cary P., Crane-Robinson C., Coles B., Goodwin G.H.;





```

[6]
RP CHROMOSOMAL TRANSLOCATION WITH RAD51L1.
RX MEDLINE=22535398; PubMed=12649198;
RA Quade B.J., Weremowicz S., Neskey D.M., Vanni R., Ladd C., Dal Cin P.,
RA Morton C.C.;
RT "Fusion transcripts involving HMGA2 are not a common molecular
RT mechanism in uterine leiomyomata with rearrangements in 12q15.";
RL Cancer Res. 63:1351-1358(2003).
CC -!- FUNCTION: Involved in transcription regulation.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: Expressed predominantly during embryogenesis.
CC -!- PTM: Regulated by cell cycle-dependent phosphorylation which
CC alters its DNA binding affinity.
CC -!- DISEASE: HMGIC is fused with a number of other genes in lipomas.
CC -!- DISEASE: Involved in uterine leiomyoma through a chromosomal
CC translocation t(12;14)(q15;q23-24) that involves HMGA2 and
CC RAD51L1.
CC -!- SIMILARITY: Contains 3 A.T hook DNA-binding repeats.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/Chromocancer/Genes/HMGICID82.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L46353; AAA96484.1; -
DR EMBL; L41044; AAA96484.1; JOINED.
DR EMBL; L44578; AAA96484.1; JOINED.
DR EMBL; Z31595; CAA83472.1; -
DR EMBL; U28754; AAA68613.1; -
DR EMBL; U28750; AAA68614.1; -
DR EMBL; U28751; AAA68614.1; JOINED.
DR EMBL; U28752; AAA68614.1; JOINED.
DR EMBL; U28753; AAA68614.1; JOINED.
DR EMBL; X22518; CAA63295.1; -
DR FIC; JC2232; JC2232.
DR TRANSFAC; T01859; -
DR Genew; HGNC:5009; HMGA2.
DR MIM; 600698; -
DR GO; GO:0003680; F:AT DNA binding; TAS.
DR GO; GO:0007275; P:development; TAS.
DR GO; GO:0006325; P:establishment and/or maintenance of chromatin...; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR000116; Highmobility_IV.
DR Pfam; PF02178; AT hook; 3.
DR PRINTS; PR00929; ATHOOK.
DR ProDom; PD005593; Highmobility_IV; 1.
DR SMART; SM00384; AT_hook; 3.
DR PROSITE; PS00354; HMG1_Y; 2.
KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat;
KW Transcription regulation; Phosphorylation; Chromosomal translocation;
KW Proto-oncogene.
FT DNA_BIND. 24 34 A.T HOOK 1.
FT DNA_BIND. 44 54 A.T HOOK 2.
FT DNA_BIND. 71 82 A.T HOOK 3.
SQ SEQUENCE 109 AA; 11832 MW; F36BABE623DA4615 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 109;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRPOG 6
Db |||||
49 RGRPKG 54

RESULT 11

```

```

RRAS HUMAN
ID RRAS_HUMAN STANDARD; PRT; 218 AA.
AC F10301;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ras-related protein R-Ras (p23).
GN RRAS
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=87078390; PubMed=3098437;
RA Lowe D.G., Capon D.J., Delwart E., Sakaguchi A.Y., Naylor S.L.,
RA Goeddel D.V.;
RT "Structure of the human and murine R-ras genes, novel genes closely
RT related to ras proto-oncogenes.";
RL Cell 48:137-146(1987).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Fuhr H.L. III, Ikeda S.R., Aronstam R.S.;
RT "cDNA clones of human proteins involved in signal transduction
RT sequenced by the Guthrie cDNA resource center (www.cdna.org).";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RC TISSUE=Brain, and Uterus;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.M., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Inner surface of plasma membrane possibly
CC with attachment requiring acylation of the C-terminal cysteine (By
CC similarity with RAS).
CC -!- SIMILARITY: Belongs to the small GTPase superfamily. Ras family..
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14949; AAA60256.1; -
DR EMBL; M14948; AAA60256.1; JOINED.
DR EMBL; AF493920; AAM12634.1; -
DR EMBL; BC016286; AAH16286.1; -
DR EMBL; BC016318; AAH16318.1; -
DR PIR; A26159; TVHUR.
DR HSPD; P01112; IFFL.
DR SWISS-2DPAGE; P10301; HUMAN.
DR Genew; HGNC:10447; RRAS.

```



```

DR MM: 165090;
DR GO: GO:0003330; F-RAS small monomeric GTPase activity; TAS.
DR GO: GO:0007265; F-RAS protein signal transduction; TAS.
DR InterPro: IPR003577; GTPase Ras.
DR InterPro: IPR001806; Ras trnsfrmg.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; ras; 1.
DR PRINTS: PR00449; RASTRNSFRMNG.
DR SMART: SM00173; RAS; 1.
DR TIGRFAMs: TIGR00231; small GTP; 1.
KW GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 36 43 GTP (BY SIMILARITY).
FT NP_BIND 83 87 GTP (BY SIMILARITY).
FT NP_BIND 142 145 GTP (BY SIMILARITY).
FT DOMAIN 58 66 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 215 215 S-geranylgeranyl cysteine
                          (By similarity).
SQ SEQUENCE 218 AA; 23480 MW; 437F73170670EB28 CRC64;

Query Match      88.2%; Score 30; DB 1; Length 218;
Best Local Similarity 83.3%; Pred. No. 37;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6
DB 11 RGRPRG 16

RESULT 12
RRAS_MOUSE
ID RRAS_MOUSE STANDARD; PRT; 218 AA.
AC P10833;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ras-related protein R-Ras (P23).
DN RRAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87078390; PubMed=3098437;
RA Lowe D.G., Capon D.J., Delwart E., Sakaguchi A.Y., Naylor S.L.,
RA Goeddel D.V.;
RT "Structure of the human and murine R-ras genes, novel genes closely
RT related to ras proto-oncogenes.";
RL Cell 48:137-146(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=23388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
RA Dratchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.E.,
RA Stapleton M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

```

---

```

CC -1- SUBCELLULAR LOCATION: Inner surface of plasma membrane possibly
CC with attachment requiring acylation of the C-terminal cysteine (By
CC similarity with RAS).
CC -1- SIMILARITY: Belongs to the small GTPase superfamily. Ras family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M21019; AAA40038.1; -.
CC EMBL; BC003105; AAH09105.1; -.
CC HSSP; P01112; IPLD.
CC MGD; MGI:98179; Kras.
CC InterPro: IPR003577; GTPase Ras.
CC InterPro: IPR001806; Ras trnsfrmg.
CC Pfam: PF00071; ras; 1.
CC PRINTS; PR00449; RASTRNSFRMNG.
CC SMART; SM00173; RAS; 1.
CC TIGRFAMs; TIGR00231; small GTP; 1.
KW GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 36 43 GTP (BY SIMILARITY).
FT NP_BIND 83 87 GTP (BY SIMILARITY).
FT NP_BIND 142 145 GTP (BY SIMILARITY).
FT DOMAIN 58 66 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 215 215 S-geranylgeranyl cysteine
                          (By similarity).
SQ SEQUENCE 218 AA; 23764 MW; CID32CE7904322B5 CRC64;

Query Match      88.2%; Score 30; DB 1; Length 218;
Best Local Similarity 83.3%; Pred. No. 37;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6
DB 11 RGRPRG 16

RESULT 13
PRP2_MOUSE
ID PRP2_MOUSE STANDARD; PRT; 261 AA.
AC P05142;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Proline-rich protein MP-2 precursor.
DN PRP1 OR PRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RX MEDLINE=86059475; PubMed=2999141;
RA Ann D.K., Carlson D.M.;
RT "The structure and organization of a proline-rich protein gene of a
RT mouse multigene family.";
RL J. Biol. Chem. 260:15863-15872(1985).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M12099; AAA40004.1; -.
CC MGD; MGI:97773; Prhl.

```

```

KW Repeat: Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 261 PROLINE-RICH PROTEIN MP-2.
SQ SEQUENCE 261 AA; 26034 MW; 36E13BA7387F47D4 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 261;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 0; Gaps 0; Indels 0;

Qy 1 RGRPOG 6
Db 254 QGRPOG 259

RESULT 14
RS31_ARATH.
ID RS31_ARATH STANDARD; PRT; 264 AA.
AC P22964; Q944A4; Q9M278;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Arginine/serine-rich splicing factor RSP31.
GN RSP31 OR AT3G61860 OR F21F14.30
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=97143875; PubMed=89989882;
Lopato S., Waigmann E., Barta A.;
RT "Characterization of a novel arginine/serine-rich splicing factor in
Arabidopsis."
RL Plant Cell 8:2255-2264(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salancubut M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Oberwäster B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,
RA Mincker P., Catolico L., Weissbach J., Surin W., Quefrier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Brandt S.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordiek G.,
RA Reichelt J., Schaefer M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwälder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Liauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA Manhaupt G., Haase D., Schöck H., Rüd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walte R., Wu D., Peterson J., Van Aken S.,
RA Creasy T.H., Haas B., Maiti R., Gill J.E., Feldblyum T.V.,
RA Pal G., Miltischer J., Sellers P., Salzborg S.L., White O., Venter J.C.,
RA Preuss D., Lin X., Nierman W.C., Salzborg S.L., Sato S., Katoh T., Asamizu E.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Katoh T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuoka A., Muraki A.,
RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana."
RL Nature 408:820-822(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;

```

```

RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.B., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Akarawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Eju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koeseema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
genome."
RL Science 302:842-846(2003).
CC -!- FUNCTION: Required for constitutive and alternative pre-mRNA
CC splicing.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: Highly expressed in roots and flowers. A
CC presumably longer alternatively spliced form is found in leaves,
CC stems and flowers.
CC -!- SIMILARITY: Belongs to the splicing factor SR family.
CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X99435; CAA67798.1; -
DR EMBL; AL138642; CAB71893.1; ALT_SEQ.
DR EMBL; AF439831; AAL27502.1; -
DR EMBL; AY125565; AAM78075.1; -
DR FIR; TS1304; TS1304.
DR HSSP; P09651; IHAL.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rtm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS00102; RRM; 2.
DR Nucleic acid binding; RNA-binding; Alternative splicing; Repeat.
FT DOMAIN 2 74 RNA-BINDING (RRM) 1.
FT DOMAIN 93 164 RNA-BINDING (RRM) 2.
FT DOMAIN 172 264 ARG/SER-RICH (RS DOMAIN).
FT CONFLICT 66 66 R -> K (IN REF. 1).
SQ SEQUENCE 264 AA; 31154 MW; 445FD2E4A5E0C203 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 264;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPOG 6
Db 76 RGRPRG 81

RESULT 15
MEC2_MOUSE
ID MEC2_MOUSE STANDARD; PRT; 484 AA.
AC Q322E6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Methyl-CpG-binding protein 2 (Mecp-2 protein) (Mecp2).
GN MECP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
CPG binding proteins.";
RL Mol. Cell. Biol. 18:6538-6547 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9929240; PubMed=10369871;
RA Coy J.F., Sedlacek Z., Baechner D., Delius H., Poustka A.;
RT "A complex pattern of evolutionary conservation and alternative
polyadenylation within the long 3'-untranslated region of the methyl-
CPG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene
expression.";
RL Hum. Mol. Genet. 8:1253-1262 (1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Reichwald K., Thiessen J., Wiehe T., Kioschis P., Straetling W.H.,
RA Rosenthal A., Platzer M.;
RT "Comparative analysis of the methyl CPG binding protein 2 locus in man
and mouse reveals new untranslated sequences.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP INTERACTION WITH FNEP3.
RX MEDLINE=97315177; PubMed=9171351;
RA Bedford M.T., Chan D.C., Leder P.;
RT "FBP WW domains and the Abl SH3 domain bind to a specific class of
proline-rich ligands.";
RL EMBO J. 16:2376-2383 (1997).
CC -!- FUNCTION: Chromosomal protein that binds to methylated DNA. It can
bind specifically to a single methyl-CpG pair. It is not
influenced by sequences flanking the methyl-CpGs. Mediates
transcriptional repression through interaction with histone
deacetylase and the corepressor SIN3A (By similarity).
CC -!- SUBUNIT: Interacts with FNEP3.
CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalized with methyl-CpG in the
genome.
CC -!- SIMILARITY: Contains 1 A.T hook DNA-binding repeat.
CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF072251; AAC6880.1; -
DR EMBL; AJ132922; CAB46495.1; -
DR EMBL; AF121351; AAF22116.1; -
DR EMBL; AF158181; AAF33024.1; -
DR MGD; MGI:99918; Mecp2.
DR GO; GO:0005634; C:nucleus; IDA.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT DOMAIN 96 149 MBD.
FT DOMAIN 277 283 POLY-ALA.
FT DOMAIN 366 372 POLY-HIS.
FT DOMAIN 384 391 POLY-PRO.
FT DOMAIN 440 443 POLY-THR.
SQ SEQUENCE 484 AA; 52307 MW; 62FD228F0118A9F CRC64;

Query Match 88.2%; Score 30; DB 1; Length 484;
Best Local Similarity 83.3%; Pred. NO. 83;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6
```

Db 188 RGRPQG 193

Search completed: April 6, 2004, 16:07:57  
Job time : 3.86916 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 21.7009 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-2

Perfect score: 34

Sequence: 1 RGRPGQ 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvrius:\*

16: sp\_bacteriaph:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	209	4 Q8N811	Q8N811 homo sapien
2	34	100.0	256	16 Q7UYG7	Q7UYG7 rhodopirell
3	34	100.0	278	2 O86059	O86059 azospirillu
4	34	100.0	422	16 Q9RKB0	Q9RKB0 streptomyc
5	34	100.0	590	16 Q7KEL9	Q7KEL9 bordetella
6	34	100.0	590	16 Q7W3A1	Q7W3A1 bordetella
7	34	100.0	590	16 Q7VU53	Q7VU53 bordetella
8	34	100.0	2188	16 P96202	P96202 mycobacteri
9	34	100.0	2188	16 Q7TXL8	Q7TXL8 mycobacteri
10	31	91.2	172	12 Q65016	Q65016 artichoke m
11	31	91.2	183	16 Q81HF7	Q81HF7 bacillus ce
12	31	91.2	195	16 Q930K0	Q930K0 rhizobium m
13	31	91.2	269	16 Q88EN4	Q88EN4 pseudomonas
14	31	91.2	278	4 Q9H5J8	Q9H5J8 homo sapien
15	31	91.2	281	16 Q880V7	Q880V7 pseudomonas
16	31	91.2	284	16 Q55556	Q55556 synechocyst

17	31	91.2	393	2 Q84J13	Q84J13 pleurocapsa
18	31	91.2	399	12 Q9QJ19	Q9QJ19 snakehead r
19	31	91.2	400	16 Q88Q76	Q88Q76 pseudomonas
20	31	91.2	429	16 Q87W63	Q87W63 pseudomonas
21	31	91.2	448	16 Q87UH9	Q87UH9 pseudomonas
22	31	91.2	450	16 Q88EJ7	Q88EJ7 pseudomonas
23	31	91.2	481	16 Q7UJ99	Q7UJ99 rhodopirell
24	31	91.2	498	15 P89686	P89686 feline immu
25	31	91.2	518	10 Q9M3X8	Q9M3X8 lens culina
26	31	91.2	625	16 Q82SH7	Q82SH7 streptomyc
27	31	91.2	655	5 Q86PB6	Q86PB6 drosophila
28	31	91.2	681	5 Q8WSS3	Q8WSS3 penaeus jap
29	31	91.2	684	5 Q8MUI3	Q8MUI3 penaeus sem
30	31	91.2	686	5 Q8MUI4	Q8MUI4 penaeus mon
31	31	91.2	688	5 Q9V1P5	Q9V1P5 penaeus mon
32	31	91.2	688	5 Q95YGI	Q95YGI penaeus jap
33	31	91.2	822	2 Q9RED8	Q9RED8 burkholderi
34	31	91.2	997	5 Q8IQL4	Q8IQL4 drosophila
35	31	91.2	1029	5 Q9VUD0	Q9VUD0 drosophila
36	31	91.2	1095	11 Q9JLQ8	Q9JLQ8 mus musculu
37	31	91.2	3590	2 Q8VV99	Q8VV99 bordetella
38	31	91.2	3590	2 Q45365	Q45365 bordetella
39	30	88.2	65	10 Q84PY6	Q84PY6 oryza sativ
40	30	88.2	77	10 Q8LI08	Q8LI08 oryza sativ
41	30	88.2	86	16 Q8FS15	Q8FS15 corynebacte
42	30	88.2	96	11 Q924I7	Q924I7 mus musculu
43	30	88.2	96	11 Q8K1F5	Q8K1F5 rattus norv
44	30	88.2	98	10 Q9AS52	Q9AS52 oryza sativ
45	30	88.2	105	11 Q9D6G0	Q9D6G0 mus musculu

## ALIGNMENTS

### RESULT 1

Q8N811 PRELIMINARY; PRT; 209 AA.

AC Q8N811;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Hypothetical protein FLJ40144.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Ohima A., Takahashi-Fujii A., Tanase T., Inose N., Takeuchi K.,  
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,  
RA Osuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,  
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK097463; BAC05064.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 209 AA; 23039 MW; DA3629D5996A2E04 CRC64;

Query Match 100.0%; Score 34; DB 4; Length 209;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RGRPGQ 6

Db 119 RGRPGQ 124

### RESULT 2

Q7UYG7

ID Q7UYG7 PRELIMINARY; PRT; 256 AA.  
 AC Q7UYG7  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN RB624.  
 OS Rhodospirillum baltica.  
 OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
 OC Planctomycetaceae; Pirellula.  
 OX NCBI\_TaxID=117;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1;  
 RX MEDLINE=22735913; PubMed=12835416;  
 RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
 RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
 RA Schlesner H., Amann R., Reinhardt R.,  
 RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
 strain 1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
 DR EMBL; BX294134; CAD71675.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 256 AA; 27003 MW; 2FAA06B5206E8B2F CRC64;  
  
 Query Match 100.0%; Score 34; DB 16; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRPQG 6  
 DB 160 RGRPQG 165  
  
 RESULT 3  
 OS86059 PRELIMINARY; PRT; 278 AA.  
 AC OS86059  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Transcription activator (fragment).  
 GN ATRAB.  
 OS Azospirillum brasilense.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
 OC Rhodospirillaceae; Azospirillum.  
 OX NCBI\_TaxID=192;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20117792; PubMed=10650197;  
 RA Revers L.F., Passaglia L.M.P., Marchal K., Frazzon J., Blaha C.G.,  
 RA Vanderleyden J., Schrank I.S.;  
 RT "Characterization of an Azospirillum brasilense Tn5 mutant with  
 enhanced N(2) fixation: the effect of CRF280 on nifh expression.";  
 RL FEMS Microbiol. Lett. 183:23-29(2000).  
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.  
 DR EMBL; AF083219; AAC33477.1; -;  
 DR GO; GO:0003700; P:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR InterPro; IPR000847; HTH\_LysR.  
 DR InterPro; IPR005119; LysR\_subst.  
 DR Pfam; PF00126; HTH 1; 1.  
 DR Pfam; PF03466; LysR\_substrate; 1.  
 DR PRINTS; PR00039; HTHLYSR.  
 DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
 DR DNA-binding; transcription; Transcription regulation.  
 KW NON TER 278  
 FT SEQUENCE 278 AA; 30931 MW; AB534587276F8E34 CRC64;  
  
 Query Match 100.0%; Score 34; DB 2; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 49;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRPQG 6  
 DB 272 RGRPQG 277  
  
 RESULT 4  
 Q9RK30 PRELIMINARY; PRT; 422 AA.  
 AC Q9RK30  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative membrane protein.  
 GN SCO3177 OR SC887.28C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke J., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Ruster S., Taylor K.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2)."  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939115; CAB59673.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 422 AA; 45608 MW; 73494FBB9C6A4832 CRC64;  
  
 Query Match 100.0%; Score 34; DB 16; Length 422;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRPQG 6  
 DB 304 RGRPQG 309  
  
 RESULT 5  
 Q7WEL9 PRELIMINARY; PRT; 590 AA.  
 AC Q7WEL9  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN BB4615.  
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=518;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RB50 / ATCC BAA-588;  
 RX MEDLINE=22827954; PubMed=12910271;  
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,  
 RA Harris D.B., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
 RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagers K.,  
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,

```

RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640451; CAE34977.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 590 AA; 65137 MW; 84746C173EA66DCC CRC64;

Query Match 100.0%; Score 34; DB 16; Length 590;
Best Local Similarity 100.0%; Pred. NO. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6
Db 439 RGRPQG 444

RESULT 6
QW3A1 PRELIMINARY; PRT; 590 AA.
AC Q7W3A1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN BPP4145.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640435; CAE39424.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 590 AA; 65137 MW; 84746C173EA66DCC CRC64;

Query Match 100.0%; Score 34; DB 16; Length 590;
Best Local Similarity 100.0%; Pred. NO. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6
Db 439 RGRPQG 444

RESULT 7
QWVU53 PRELIMINARY; PRT; 590 AA.
AC Q7WVU53;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN BP1276.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

```

```

OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640421; CAE43542.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 590 AA; 65163 MW; F7B0F1403F3C45A9 CRC64;

Query Match 100.0%; Score 34; DB 16; Length 590;
Best Local Similarity 100.0%; Pred. NO. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6
Db 439 RGRPQG 444

RESULT 8
P96202 PRELIMINARY; PRT; 2188 AA.
AC P96202;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PPSC (Polyketide synthase).
GN PPSC OR RV2933 OR MT3003 OR MTCV19H9.01 OR MTV01.1.02.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z83857; CAB06099.1; -.
DR EMBL; AE007122; AAK47330.1; -.

```

DR PIR; A70984; A70984.  
 DR TIGR; MT3003; -.  
 DR TubercuList; Rv2933; -.  
 DR GO; GO:0004024; F1alcohol dehydrogenase activity, zinc-dependent; IEA.  
 DR GO; GO:0016740; F1transferase activity; IEA.  
 DR GO; GO:0008270; F1zinc ion binding; IEA.  
 DR GO; GO:0006633; F1fatty acid biosynthesis; IEA.  
 DR GO; GO:0008152; P1metabolism; IEA.  
 DR InterPro; IPR001227; Ac trans.  
 DR InterPro; IPR002085; Adh\_zn\_family.  
 DR InterPro; IPR000794; Ketoacyl\_synth.  
 DR InterPro; IPR006162; Ppantng\_S.  
 DR InterPro; IPR006163; Pp bind.  
 DR Pfam; PF00698; AChL\_transf; 1.  
 DR Pfam; PF00107; ADH\_zinc\_N; 1.  
 DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 1.  
 DR Pfam; PF00550; Pp-binding; 1.  
 DR PROSITE; PS00075; ACP\_DOMAIN; 1.  
 DR PROSITE; PS00606; B\_KETOACYL SYNTHASE; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 KW Phosphopantetheine; Transferase; Complete proteome.  
 FT CONFLICT 864 864 S -> R (IN REF. 2).  
 SQ SEQUENCE 2188 AA; 230619 MW; 3C431C011F01F1A2 CRC64;  
 Query Match 100.0%; Score 34; DB 16; Length 2188;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
 |||||  
 Db 1479 RGRPOG 1484

RESULT 9  
 Q7TXLS PRELIMINARY; PRT; 2188 AA.  
 AC Q7TXLS;  
 DT 01-OCT-2003 (TRENBLrel. 25, Created)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Phenolphthiocerol synthesis type-I polyketide synthase PPSC.  
 GN PPSC OR MB2958.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AF2122/57;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,  
 RA Fryor M., Dutfoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Farthill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,  
 RA "The complete genome sequence of Mycobacterium bovis";  
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).  
 DR EMBL; BX248344; CAD96645.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 2188 AA; 230620 MW; 3C431C011F01F1A2 CRC64;  
 Query Match 100.0%; Score 34; DB 16; Length 2188;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
 |||||  
 Db 1479 RGRPOG 1484

RESULT 10  
 Q65016 PRELIMINARY; PRT; 172 AA.  
 ID Q65016

AC Q65016;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Protein of 19 kDa.  
 OS Artichoke mottled crinkle virus (AMCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
 OC Tombusvirus.  
 OX NCBI\_TaxID=12142;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AMCV-Bari Dr.Gallitelli isolate;  
 RX MEDLINE=94292901; PubMed=8021582;  
 RA Tavazza M., Luciolli A., Calogero A., Pay A., Tavazza R.;  
 RT "Nucleotide sequence, genomic organization and synthesis of infection  
 RT transcripts from full-length clone of artichoke mottle crinkle  
 RT virus."; J. Gen. Virol. 75:1515-1524 (1994).  
 RL J. Gen. Virol. 75:1515-1524 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AMCV-Bari Dr.Gallitelli isolate;  
 RA Tavazza M.;  
 RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X62493; CAA44359.1; -.  
 DR GO; GO:0019012; C:virion; IEA.  
 DR InterPro; IPR004905; Tombus\_P19.  
 DR Pfam; PF03220; Tombus\_P19; 1.  
 DR ProDom; PD005743; Tombus\_P19; 1.  
 SQ SEQUENCE 172 AA; 19320 MW; 7C9C674E6DB1DB5C CRC64;  
 Query Match 91.2%; Score 31; DB 12; Length 172;  
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
 |||||  
 Db 158 RGRPEG 163

RESULT 11  
 Q81HF7 PRELIMINARY; PRT; 183 AA.  
 ID Q81HF7;  
 AC Q81HF7;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN EC0857.  
 OS Bacillus cereus (strain ATCC 14579 / DSM 31).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=226900;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22608415; PubMed=12721630;  
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,  
 RA Kapral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,  
 RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,  
 RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,  
 RA Overbeek R., Kyrpides N.;  
 RT "Genome sequence of Bacillus cereus and comparative analysis with  
 RT Bacillus anthracis";  
 RL Nature 423:87-91 (2003).  
 DR EMBL; AE017000; AAP07844.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 183 AA; 20645 MW; 0909095BF426BAC4 CRC64;

QY 1 RGRPOG 6  
 |||||  
 Db 94 RGRPOG 99

Query Match 91.2%; Score 31; DB 16; Length 183;  
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
RT metabolically versatile Pseudomonas putida KT2440."
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL; AE016790; AAN69997.1; -.
DR TIGR; PP4420; -.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; rve; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 269 AA; 31473 MW; 77BA5C83D80F7CCD CRC64;

Query Match          91.2%; Score 31; DB 16; Length 269;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6
   |||||
Db 160 RGRPQG 165

RESULT 14
Q880V7 PRELIMINARY; PRT; 278 AA.
ID Q880V7
AC Q880V7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ23363.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Oktani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.,
RA "NEDO human cDNA sequencing project."
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027016; BAB15628.1; -.
DR EMBL; BC001972; AAH01972.1; -.
KW Hypothetical protein.
SQ SEQUENCE 278 AA; 32058 MW; 4BA37C73BB387446 CRC64;

Query Match          91.2%; Score 31; DB 4; Length 278;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6
   |||||
Db 106 RGRPEG 111

RESULT 15
Q880V7 PRELIMINARY; PRT; 281 AA.
ID Q880V7
AC Q880V7
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ISPay12, transposase OrfB.
GN PSPT0304
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.

RT
```



RC STRAIN=DC3000;  
RA Buell R., Jcardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,  
RA Dodson K., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidse T.,  
RA White O., Fraser C., Collmer A.;  
RT "Complete sequence of Pseudomonas syringae."  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016866; AA056533.1; --  
DR TIGR; PSP03044; --  
DR GO; GO:0003677; P:DNA binding; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR InterPro; IPR001584; Rve.  
DR Pfam; PF00665; rve; 1.  
KW Complete proteome.  
SQ SEQUENCE 281 AA; 32887 MW; E8DA6ABE52F9ED98 CRC64;

Query Match 91.2%; Score 31; DB 16; Length 281;  
Best Local Similarity 83.3%; Pred. NO. 2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 RGRPOG 6  
||:|  
Db 172 RGRPOG 177

Search completed: April 6, 2004, 16:14:37  
Job time : 21.7779 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 33.5888 seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-2  
Perfect score: 34  
Sequence: 1 RGRPQ 6

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_29Jan04.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	6	AB48772	Human sal
2	34	100.0	7	AB48773	Human sal
3	34	100.0	8	AB48774	Human sal
4	34	100.0	9	AB48775	Human sal
5	34	100.0	10	AB48776	Human sal
6	34	100.0	17	AB48783	Human sal
7	34	100.0	37	AM14754	Peptide #
8	34	100.0	37	AB33719	Peptide #
9	34	100.0	37	AB27177	Peptide #
10	34	100.0	37	AB28535	Peptide #
11	34	100.0	37	AB19166	Protein #
12	34	100.0	37	AM66893	Human bon
13	34	100.0	37	AM54488	Human bra
14	34	100.0	37	AB48557	Human liv
15	34	100.0	37	AM02477	Peptide #
16	34	100.0	37	AB36549	Human pep
17	34	100.0	132	AB38848	Peptide #
18	34	100.0	132	AM32323	Peptide #
19	34	100.0	132	AM72058	Human bon
20	34	100.0	132	AM59494	Human bra
21	34	100.0	132	AB53742	Human liv
22	34	100.0	132	AB541873	Human pep
23	34	100.0	149	AB57423	Human NOV
24	34	100.0	154	AB56769	Human sec
25	34	100.0	166	ADA83798	Human PRH

ALIGNMENTS

RESULT 1

AB48772  
ID AB48772 standard; peptide; 6 AA.

XX AC AAB48772;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 106-111), SEQ ID NO:2.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;

XX KW Chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
XX KW oral bacterium; caries prevention.

XX OS Homo sapiens.

XX PN WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SE000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX PA (STRO/) STROEMBERG N.  
XX PA (JOHA/) JOHANSSON I.

XX PI Stroemberg N, Johansson I;

XX DR WPI; 2001-031923/04.

XX PT New oligopeptides comprising 2 arginine residues from degradation of  
XX PT proline-rich proteins, useful for preventing dental caries.

XX PS Claim 4; Page 24; 36pp; English.

XX CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
XX CC AAB48783) which contain at least two arginine residues and which protect  
XX CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
XX CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
XX CC determinants of a person's susceptibility to dental caries. PRPs are  
XX CC degraded by Actinomyces and Streptococcus species to small peptide  
XX CC fragments. These are metabolised by oral bacteria for nutritional  
XX CC purposes, with certain bacterial species generating ammonia via the  
XX CC catabolism of arginine. The peptides of the invention, being arginine-  
XX CC rich, can also be converted to ammonia by these bacteria. The ammonia  
XX CC thus formed raises the pH at the dental surface, thereby protecting the  
XX CC teeth against caries. Sequences AAB48771-548783 represent the PRP-1-

App. cont

CC derived oligopeptides of the invention

XX Sequence 6 AA;

Query Match 100.0%; Score 34; DB 4; Length 6;

Best Local Similarity 100.0%; Pred. No. 1.4e+06; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6

Db 1 RGRPQG 6

RESULT 2

ID AAB48773 standard; peptide; 7 AA.

XX AAB48773;

DT 09-MAR-2001 (first entry)

DE Human saliva PRP-1 fragment (residues 106-112), SEQ ID NO:3.

KW Human; PRP-1; proline-rich protein; saliva; dental caries; chromosome 12p13.2; arginine catabolism; ammonia production; pH increase; oral bacterium; caries prevention.

OS Homo sapiens.

FN WO200069890-A1.

PD 23-NOV-2000.

PF 11-MAY-2000; 2000WO-SE000930.

PR 17-MAY-1999; 99SE-00001773.

PA (STRO/) STROEMBERG N.

PI (JOHA/) JOHANSSON I.

PI Stroemberg N, Johansson I;

DR WPI; 2001-031923/04.

PT New oligopeptides comprising 2 arginine residues from degradation of proline-rich proteins, useful for preventing dental caries.

PS Claim 4; Page 24; 36pp; English.

CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-AAB48783) which contain at least two arginine residues and which protect against dental caries. PRPs (proline-rich proteins) are salivary proteins encoded by six clustered genes on chromosome 12p13.2 and are potential determinants of a person's susceptibility to dental caries. PRPs are degraded by Actinomyces and Streptococcus species to small peptide fragments. These are metabolised by oral bacteria for nutritional purposes, with certain bacterial species generating ammonia via the catabolism of arginine. The peptides of the invention, being arginine-rich, can also be converted to ammonia by these bacteria. The ammonia thus formed raises the pH at the dental surface, thereby protecting the teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-derived oligopeptides of the invention

XX Sequence 7 AA;

Query Match

Best Local Similarity 100.0%; Score 34; DB 4; Length 7;

Mismatches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6

Db 1 RGRPQG 6

RESULT 3

ID AAB48774 standard; peptide; 8 AA.

XX AAB48774;

DT 09-MAR-2001 (first entry)

DE Human saliva PRP-1 fragment (residues 106-113), SEQ ID NO:4.

KW Human; PRP-1; proline-rich protein; saliva; dental caries; chromosome 12p13.2; arginine catabolism; ammonia production; pH increase; oral bacterium; caries prevention.

OS Homo sapiens.

FN WO200069890-A1.

PD 23-NOV-2000.

PF 11-MAY-2000; 2000WO-SE000930.

PR 17-MAY-1999; 99SE-00001773.

PA (STRO/) STROEMBERG N.

PI (JOHA/) JOHANSSON I.

PI Stroemberg N, Johansson I;

DR WPI; 2001-031923/04.

PT New oligopeptides comprising 2 arginine residues from degradation of proline-rich proteins, useful for preventing dental caries.

PS Claim 4; Page 24; 36pp; English.

CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-AAB48783) which contain at least two arginine residues and which protect against dental caries. PRPs (proline-rich proteins) are salivary proteins encoded by six clustered genes on chromosome 12p13.2 and are potential determinants of a person's susceptibility to dental caries. PRPs are degraded by Actinomyces and Streptococcus species to small peptide fragments. These are metabolised by oral bacteria for nutritional purposes, with certain bacterial species generating ammonia via the catabolism of arginine. The peptides of the invention, being arginine-rich, can also be converted to ammonia by these bacteria. The ammonia thus formed raises the pH at the dental surface, thereby protecting the teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-derived oligopeptides of the invention

XX Sequence 8 AA;

Query Match

Best Local Similarity 100.0%; Score 34; DB 4; Length 8;

Mismatches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6

Db 1 RGRPQG 6

RESULT 4

ID AAB48775 standard; peptide; 9 AA.

XX AAB48775;

DT 09-MAR-2001 (first entry)

DE Human saliva PRP-1 fragment (residues 106-114), SEQ ID NO:5.

KW Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX Homo sapiens.  
 OS WO200069890-A1.  
 FN 23-NOV-2000.  
 XX 11-MAY-2000; 2000WO-SE000930.  
 PF 17-MAY-1999; 99SE-00001773.  
 PR (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX Stroemberg N, Johansson I;  
 XX WPI; 2001-031923/04.  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX Claim 4; Page 24; 36pp; English.  
 XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX Sequence 9 AA;  
 SQ Query Match 100.0%; Score 34; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPQG 6  
 DB |||||  
 1 RGRPQG 6  
 RESULT 5  
 AAB48776  
 ID AAB48776 standard; peptide; 10 AA.  
 XX AAB48776;  
 AC 09-MAR-2001 (first entry)  
 DT Human saliva PRP-1 fragment (residues 106-115), SEQ ID NO:6.  
 DE Human; PRP-1; proline-rich protein; saliva; dental caries;  
 XX chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX Homo sapiens.  
 OS WO200069890-A1.  
 FN 23-NOV-2000.  
 XX 11-MAY-2000; 2000WO-SE000930.  
 PF 17-MAY-1999; 99SE-00001773.  
 PR (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX Stroemberg N, Johansson I;  
 XX WPI; 2001-031923/04.  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX Claim 2; Page 24; 36pp; English.

XX (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX Stroemberg N, Johansson I;  
 XX WPI; 2001-031923/04.  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX Claim 4; Page 24; 36pp; English.  
 XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX Sequence 10 AA;  
 SQ Query Match 100.0%; Score 34; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 6.3;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPQG 6  
 DB |||||  
 1 RGRPQG 6  
 RESULT 6  
 AAB48783  
 ID AAB48783 standard; peptide; 17 AA.  
 XX AAB48783;  
 AC 09-MAR-2001 (first entry)  
 DT Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.  
 DE Human; PRP-1; proline-rich protein; saliva; dental caries;  
 KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX Homo sapiens.  
 OS WO200069890-A1.  
 FN 23-NOV-2000.  
 XX 11-MAY-2000; 2000WO-SE000930.  
 PF 17-MAY-1999; 99SE-00001773.  
 PR (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX Stroemberg N, Johansson I;  
 XX WPI; 2001-031923/04.  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX Claim 2; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention

XX Sequence 17 AA;

Query Match 100.0%; Score 34; DB 4; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
 |||||  
 DB 8 RGRPOG 13

#### RESULT 7

AA014754  
 ID AAM14754 standard; protein; 37 AA.

XX AC AAM14754;

XX DT 12-OCT-2001 (first entry)

XX DE Peptide #1188 encoded by probe for measuring cervical gene expression.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 CC cervical cancer.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000670.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 CC gene expression in human cervical epithelial cells.

XX PS Claim 27; SEQ ID NO 19580; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
 CC (SNP; see A110068-A128459). The present sequence is a peptide encoded  
 CC by one such probe. The SNPs are derived from human Hela cells. The SNPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably

CC cervical cancer. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 37 AA;

Query Match 100.0%; Score 34; DB 4; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
 |||||  
 DB 13 RGRPOG 18

#### RESULT 8

ABB33719  
 ID ABB33719 standard; peptide; 37 AA.

XX AC ABB33719;

XX DT 04-FEB-2002 (first entry)

XX DE Peptide #1225 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000669.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 CC gene expression in human fetal liver.

XX PS Claim 27; SEQ ID NO 26354; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring  
 CC human gene expression in a sample derived from human foetal liver. The  
 CC single exon nucleic acid probes may be used for predicting, measuring and  
 CC displaying gene expression in samples derived from human fetal liver. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid probe  
 CC of the invention. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 37 AA;

Query Match 100.0%; Score 34; DB 4; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
 |||||  
 DB 13 RGRPOG 18

```

RESULT 9
ID AM27177 standard; protein; 37 AA.
XX
XX AM27177;
XX
XX 17-OCT-2001 (first entry)
XX
XX Peptide #1114 encoded by probe for measuring placental gene expression.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX Claim 27; SEQ ID NO 27446; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP;
XX see AA13135-AA15746). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
XX Sequence 37 AA;
XX
XX Query Match 100.0%; Score 34; DB 4; Length 37;
XX Best Local Similarity 100.0%; Pred. NO. 22;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RGRPQG 6
XX |||||
XX 13 RGRPQG 18
XX
XX RESULT 10
XX ABB28535
XX ID ABB28535 standard; peptide; 37 AA.
XX
XX AC ABB28535;
XX
XX XX 01-FEB-2002 (first entry)
XX
XX Peptide #1186 encoded by breast cell single exon nucleic acid probe.
XX
XX Human; microarray; single exon probe; gene expression; breast; disease;
XX cancer.
XX
XX Homo sapiens.
XX

```

```

XX WO200157271-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000662.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes, useful
XX for measuring gene expression in sample derived from human breast,
XX comprises number of single exon nucleic acid probes.
XX
XX Claim 27; SEQ ID NO 11503; 327pp + Sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and Br 474 cells. The method involves contacting the
XX probes with a collection of detectably labelled nucleic acids derived
XX from mRNA of human breast, and then measuring the label bound to each
XX probe of the microarray. The probes are useful for verifying the
XX expression of regions of genomic DNA predicted to encode proteins. They
XX are useful for gene discovery, and for determining predisposition and/or
XX prognosing breast disease. Gene expression analysis is useful for
XX assessing the toxicity of chemical agents on cells. The microarray of
XX this invention presents a far greater diversity of probes for measuring
XX gene expression, with far less bias than expressed sequence tag
XX microarrays. The method is suitable for rapid production of functional
XX information from genomic sequence. The present sequence is a peptide
XX encoded by a single exon nucleic acid probe of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 37 AA;
XX
XX Query Match 100.0%; Score 34; DB 4; Length 37;
XX Best Local Similarity 100.0%; Pred. NO. 22;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RGRPQG 6
XX |||||
XX 13 RGRPQG 18
XX
XX Db
XX
XX RESULT 11
XX ABB19166
XX ID ABB19166 standard; protein; 37 AA.
XX
XX AC ABB19166;
XX
XX XX 23-JAN-2002 (first entry)
XX
XX Protein #1165 encoded by probe for measuring heart cell gene expression.
XX
XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX
XX Homo sapiens.
XX

```

```

PN WO200157274-A2.
XX
PD
PR 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000666.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488899/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 20936; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21533-AAA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 37 AA;
XX
XX
XX Query Match 100.0%; Score 34; DB 4; Length 37;
XX Best Local Similarity 100.0%; Pred. No. 22;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RGRPQG 6
XX |||||
XX 13 RGRPQG 18
XX
XX
XX RESULT 12
XX AAM66893
XX ID AAM66893 standard; protein; 37 AA.
XX
XX AC AAM66893;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27199.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000668.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
XX
XX Example 4; SEQ ID NO 26593; 650pp + Sequence Listing; English.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
XX Claim 15; SEQ ID NO 20936; 530pp; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX
XX Sequence 37 AA;
XX
XX
XX Query Match 100.0%; Score 34; DB 4; Length 37;
XX Best Local Similarity 100.0%; Pred. No. 22;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RGRPQG 6
XX |||||
XX 13 RGRPQG 18
XX
XX
XX RESULT 13
XX AAM54488
XX ID AAM54488 standard; protein; 37 AA.
XX
XX AC AAM54488;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26593.
XX
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
XX OS Homo sapiens.
XX
XX PN WO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO 26593; 650pp + Sequence Listing; English.
XX

```

XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention

XX  
 SQ Sequence 37 AA;  
 Query Match 100.0%; Score 34; DB 4; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6  
 DB 13 RGRPQG 18  
 |||||  
 |||||

RESULT 14  
 ABG48557  
 ID ABG48557 standard; peptide; 37 AA.  
 AC ABG48557;  
 XX  
 DT 25-FEB-2003 (first entry)  
 XX  
 DE Human liver peptide, SEQ ID NO 27205.  
 XX  
 KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157273-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000664.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488898/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human adult liver.  
 XX  
 PS Claim 27; SEQ ID NO 27205; 658pp; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABG47348-ABG5930 represent human  
 CC liver single exon encoded peptides of the invention. Note: The sequence  
 CC information for this patent does not appear in the printed specification  
 CC but was obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 37 AA;  
 SQ  
 Query Match 100.0%; Score 34; DB 4; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6  
 DB 13 RGRPQG 18  
 |||||  
 |||||

RESULT 15  
 AAM02477  
 ID AAM02477 standard; protein; 37 AA.  
 XX  
 AC AAM02477;  
 XX  
 DT 09-OCT-2001 (first entry)  
 XX  
 DE Peptide #1159 encoded by probe for measuring breast gene expression.  
 XX  
 KW Probe; human; breast disease; breast cancer; development disorder;  
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157270-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 29-JAN-2001; 2001WO-US000661.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-476286/51.  
 XX  
 PT Novel single exon nucleic acid probe used to measuring gene expression in  
 PT a human breast.  
 XX  
 PS Claim 27; SEQ ID NO 11217; 322pp; English.  
 XX  
 CC The present invention relates to novel single exon nucleic acid probes  
 CC (see AAI00010-AAI10067). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for measuring human gene expression in  
 CC a human breast sample, where the probe hybridises at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
 CC breast disease and non-carcinoma tumours. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 37 AA;  
 Query Match 100.0%; Score 34; DB 4; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 RGRPQG 6  
|||  
Db 13 RGRPQG 18

Search completed: April 6, 2004, 16:06:36  
Job time : 34.5888 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 23.5514 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-2  
Perfect score: 34  
Sequence: 1 RGRPQG 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	37	9	US-09-864-761-34464
2	34	100.0	89	12	US-10-424-599-195980
3	34	100.0	132	9	US-09-864-761-43644
4	34	100.0	163	12	US-10-424-599-189117
5	34	100.0	166	14	US-10-157-031-80
6	34	100.0	209	15	US-10-108-260A-4370
7	34	100.0	332	12	US-10-425-114-59820
8	34	100.0	1283	12	US-10-282-122A-64972
9	34	100.0	1832	12	US-10-282-122A-62735
10	31	91.2	147	11	US-10-425-114-50858
11	31	91.2	174	11	US-09-864-408A-6896
12	31	91.2	217	9	US-09-925-301-1073
13	31	91.2	278	10	US-09-890-688-156
14	31	91.2	278	10	US-10-229-345-20
15	31	91.2	278	12	US-10-274-177-20

16	31	91.2	278	15	US-10-341-434-18	Sequence 18, Appl
17	31	91.2	278	15	US-10-087-080-26	Sequence 26, Appl
18	31	91.2	278	15	US-10-087-080-26	Sequence 26, Appl
19	31	91.2	477	12	US-10-425-114-43312	Sequence 43312, A
20	31	91.2	498	9	US-09-946-239-8	Sequence 8, Appl
21	31	91.2	550	12	US-10-425-114-51286	Sequence 51286, A
22	31	91.2	600	12	US-10-282-122A-50005	Sequence 50005, A
23	31	91.2	625	14	US-10-156-761-15008	Sequence 15008, A
24	30	88.2	10	10	US-09-572-404B-1576	Sequence 1576, Ap
25	30	88.2	11	9	US-09-953-587-32	Sequence 32, Appl
26	30	88.2	11	9	US-09-839-743-18	Sequence 18, Appl
27	30	88.2	11	10	US-09-226-402-32	Sequence 32, Appl
28	30	88.2	45	12	US-10-424-599-249209	Sequence 249209,
29	30	88.2	48	12	US-10-424-599-152056	Sequence 152056,
30	30	88.2	63	12	US-10-424-599-267427	Sequence 267427,
31	30	88.2	72	12	US-10-424-599-194241	Sequence 194241,
32	30	88.2	74	12	US-10-424-599-228360	Sequence 228360,
33	30	88.2	83	9	US-09-864-761-41234	Sequence 41234, A
34	30	88.2	85	14	US-10-106-698-5621	Sequence 5621, Ap
35	30	88.2	94	11	US-09-864-408A-2628	Sequence 2628, Ap
36	30	88.2	97	12	US-10-425-114-48775	Sequence 48775, A
37	30	88.2	100	12	US-10-425-114-42101	Sequence 42101, A
38	30	88.2	106	12	US-10-424-599-249497	Sequence 249497,
39	30	88.2	148	12	US-10-425-114-38432	Sequence 38432, A
40	30	88.2	163	12	US-10-425-114-53125	Sequence 53125, A
41	30	88.2	167	12	US-10-424-599-164389	Sequence 164389,
42	30	88.2	174	12	US-10-424-599-162892	Sequence 162892,
43	30	88.2	199	12	US-10-425-114-56750	Sequence 56750, A
44	30	88.2	218	10	US-09-873-546-14	Sequence 14, Appl
45	30	88.2	218	13	US-10-067-813-17	Sequence 17, Appl

## ALIGNMENTS

## RESULT 1

```
US-09-864-761-34464
; Sequence 34464, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 34464  
LENGTH: 37  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC007179.3  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.85  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.88  
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.2  
OTHER INFORMATION: EST HUMAN HIT: BF310193.1, EVALUE 7.40e+00  
OTHER INFORMATION: SWISSPROT HIT: P21573, EVALUE 4.70e+00  
US-09-864-761-34464

Query Match 100.0%; Score 34; DB 9; Length 37;

Best Local Similarity 100.0%; Pred. No. 26; Mismatches 0; Indels 0; Gaps 0;  
Matches 6; Conservative 0;

QY 1 RGRPQG 6  
|||||  
DB 13 RGRPQG 18

## RESULT 2

US-10-424-599-195980  
Sequence 195980, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovacic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 195980  
LENGTH: 89  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_18997C.1.pep  
US-10-424-599-195980

Query Match 100.0%; Score 34; DB 12; Length 89;

Best Local Similarity 100.0%; Pred. No. 58;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6  
|||||  
DB 22 RGRPQG 27

## RESULT 3

US-09-864-761-43644  
Sequence 43644, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aecmca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 43644  
LENGTH: 132  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC006518.17  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77  
OTHER INFORMATION: EST HUMAN HIT: BF088785.1, EVALUE 1.00e-01  
OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02  
US-09-864-761-43644

Query Match 100.0%; Score 34; DB 9; Length 132;

Best Local Similarity 100.0%; Pred. No. 84;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
|  
|  
|  
|  
|  
Db 88 RGRPOG 93

## RESULT 4

US-10-424-599-189117  
; Sequence 189117, Application US/10424599  
; Publication No. US20040031072A1

## GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 189117

; LENGTH: 163

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_141788C.1.pep

US-10-424-599-189117

Query Match 100.0%; Score 34; DB 12; Length 163;

Best Local Similarity 100.0%; Pred. No. 1e+02; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6

|  
|  
|  
|  
|  
Db 139 RGRPOG 144

## RESULT 5

US-10-157-031-80

; Sequence 80, Application US/10157031

; Publication No. US20030108890A1

## GENERAL INFORMATION:

; APPLICANT: Baranova, A. V.

; APPLICANT: Yankovsky, N. K.

; APPLICANT: Kozlov, A. P.

; APPLICANT: Lobashev, A. V.

; APPLICANT: Krukovskaya, L. L.

; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences

; FILE REFERENCE: 2760-103

; CURRENT APPLICATION NUMBER: US/10/157,031

; CURRENT FILING DATE: 2002-05-30

; NUMBER OF SEQ ID NOS: 415

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 80

; LENGTH: 166

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-157-031-80

Query Match 100.0%; Score 34; DB 14; Length 166;

Best Local Similarity 100.0%; Pred. No. 1e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6

|  
|  
|  
|  
|  
Db 122 RGRPOG 127

## RESULT 6

US-10-108-260A-4370

; Sequence 4370, Application US/10108260A

; Publication No. US20040005560A1

## GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE

; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA

; FILE REFERENCE: H1-A0106

; CURRENT APPLICATION NUMBER: US/10/108,260A

; CURRENT FILING DATE: 2002-03-27

; NUMBER OF SEQ ID NOS: 5458

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4370

; LENGTH: 209

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-108-260A-4370

Query Match 100.0%; Score 34; DB 15; Length 209;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6

|  
|  
|  
|  
|  
Db 119 RGRPOG 124

## RESULT 7

US-10-425-114-59820

; Sequence 59820, Application US/10425114

; Publication No. US20040034888A1

## GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 59820

; LENGTH: 332

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3912-044-H8\_FLI.pep

US-10-425-114-59820

Query Match 100.0%; Score 34; DB 12; Length 332;

Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6

|  
|  
|  
|  
|  
Db 203 RGRPOG 208

## RESULT 8

US-10-282-122A-64972

; Sequence 64972, Application US/10282122A

; Publication No. US20040029129A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

```

; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64972
; LENGTH: 1283
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; NAME/KEY: MISC FEATURE
; LOCATION: (696)..(696)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (698)..(698)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-64972

Query Match      100.0%; Score 34; DB 12; Length 1283;
Best Local Similarity 100.0%; Pred. No. 7.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRPOG 6
Db      577 RGRPOG 582

RESULT 9
US-10-282-122A-62735
; Sequence 62735, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078

```

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62735
; LENGTH: 1832
; TYPE: PRT
; ORGANISM: Mycobacterium bovis
US-10-282-122A-62735

Query Match      100.0%; Score 34; DB 12; Length 1832;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRPOG 6
Db      1123 RGRPOG 1128

RESULT 10
US-10-425-114-50858
; Sequence 50858, Application US/10425114
; Publication No. US20040034688A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 50858
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LTB3075-012-F7_FJI.psp
US-10-425-114-50858

Query Match      91.2%; Score 31; DB 12; Length 147;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRPOG 6
Db      83 RGRPEG 88

RESULT 11
US-09-864-408A-6896

```

```

; Sequence 6896, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1el Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6896
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-6896

Query Match          91.2%; Score 31; DB 11; Length 174;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6
DB 44 RGRPEG 49

RESULT 12
US-09-925-301-1073
; Sequence 1073, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1073
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; FEATURE:
; LOCATION: (68)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1073

Query Match          91.2%; Score 31; DB 9; Length 217;
Best Local Similarity 83.3%; Pred. No. 4.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6
DB 111 RGRPEG 116

RESULT 13
US-09-890-688-156
; Sequence 156, Application US/09890688
; Publication No. US20030144475A1
; GENERAL INFORMATION:
; APPLICANT: Seishi KATO
; APPLICANT: Chikashi EGUCHI
; APPLICANT: Mihoro SAEKI
; TITLE OF INVENTION: Human Proteins and cDNAs thereof
; FILE REFERENCE: 2001-1102A/WMC/00653

```

```

; CURRENT APPLICATION NUMBER: US/09/890,688
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: JP 11-346863
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: JP 11-34684
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: JP 2000-31062
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: JP 2000-34091
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: JP 2000-34090
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: JP 2000-35829
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 2000-35899
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 2000-71161
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: JP 2000-160851
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 156
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-890-688-156

Query Match          91.2%; Score 31; DB 10; Length 278;
Best Local Similarity 83.3%; Pred. No. 6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6
DB 106 RGRPEG 111

RESULT 14
US-10-229-345-20
; Sequence 20, Application US/10229345
; Publication No. US20040038220A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/229,345
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-345-20

Query Match          91.2%; Score 31; DB 12; Length 278;
Best Local Similarity 83.3%; Pred. No. 6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6
DB 106 RGRPEG 111

RESULT 15
US-10-274-177-20
; Sequence 20, Application US/10274177
; Publication No. US20040038225A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/274,177

```

```
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US/10/229,345
; PRIOR FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO '20
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-274-177-20

Query Match      91.2%; Score 31; DB 12; Length 278;
Best Local Similarity 83.3%; Pred. No. 6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRPOG 6
Db      106 RGRPEG 111

Search completed: April 6, 2004, 17:05:51
Job time : 24.5514 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 8.80374 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-2

Sequence: 34

Scoring table: 1 RGRPOG 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

#### Database :

Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	392	4	US-09-252-991A-27798
2	34	100.0	1646	4	US-09-252-991A-22312
3	31	91.2	133	4	US-09-252-991A-18533
4	31	91.2	217	4	US-09-252-991A-22224
5	31	91.2	291	4	US-09-252-991A-19638
6	31	91.2	423	4	US-09-489-039A-7834
7	31	91.2	498	3	US-09-238-303-8
8	31	91.2	498	4	US-09-946-239-8
9	31	91.2	787	4	US-09-252-991A-19991
10	31	91.2	894	4	US-09-252-991A-30397
11	31	91.2	1891	2	US-08-804-227C-12
12	31	91.2	1891	2	US-08-804-198-6
13	30	88.2	11	4	US-09-839-743-18
14	30	88.2	96	3	US-09-258-373-22
15	30	88.2	109	4	US-09-341-444A-2
16	30	88.2	114	4	US-09-252-991A-25597
17	30	88.2	138	4	US-09-252-991A-20111
18	30	88.2	143	3	US-09-258-373-3
19	30	88.2	210	4	US-09-053-374A-7
20	30	88.2	218	1	US-08-247-946A-3
21	30	88.2	218	5	PCT-US95-06420-3
22	30	88.2	305	4	US-09-252-991A-28580
23	30	88.2	343	4	US-08-894-454-161
24	30	88.2	343	4	US-09-252-991A-29455
25	30	88.2	408	4	US-09-252-991A-20095
26	30	88.2	446	4	US-09-252-991A-27110
27	30	88.2	454	4	US-09-489-039A-7610

```

28 30 88.2 541 4 US-09-252-991A-19837 Sequence 19837, A
29 30 88.2 633 4 US-09-252-991A-26576 Sequence 26576, A
30 30 88.2 710 4 US-09-252-991A-22478 Sequence 22478, A
31 30 88.2 891 4 US-09-252-991A-30624 Sequence 30624, A
32 30 88.2 2265 2 US-08-149-097D-36 Sequence 36, Appl
33 30 88.2 2509 2 US-08-149-097D-35 Sequence 35, Appl
34 29 85.3 29 2 US-08-896-176-4 Sequence 4, Appl
35 29 85.3 48 2 US-08-896-176-1 Sequence 1, Appl
36 29 85.3 58 4 US-09-304-615-157 Sequence 157, App
37 29 85.3 64 4 US-09-370-838-74 Sequence 74, Appl
38 29 85.3 73 1 US-08-355-888A-11 Sequence 11, Appl
39 29 85.3 73 2 US-08-693-697-11 Sequence 11, Appl
40 29 85.3 73 3 US-08-693-696-11 Sequence 11, Appl
41 29 85.3 73 4 US-09-357-914-11 Sequence 11, Appl
42 29 85.3 81 4 US-09-415-099-4 Sequence 4, Appl
43 29 85.3 114 4 US-09-199-637A-257 Sequence 257, App
44 29 85.3 132 4 US-09-252-991A-22177 Sequence 22177, A
45 29 85.3 147 4 US-09-252-991A-23687 Sequence 23687, A

```

#### ALIGNMENTS

##### RESULT 1

```

US-09-252-991A-27798
; Sequence 27798, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27798
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27798

```

```

Query Match      100.0%; Score 34; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. NO. 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 RGRPOG 6

DB 114 RGRPOG 119

##### RESULT 2

```

US-09-252-991A-22312
; Sequence 22312, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22312
; LENGTH: 1646
; TYPE: PRT

```



ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22312

Query Match 100.0%; Score 34; DB 4; Length 1646;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
| | | | |  
Db 58 RGRPQG 63

## RESULT 3

US-09-252-991A-18533  
; Sequence 18533, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18533  
; LENGTH: 133  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18533

Query Match 91.2%; Score 31; DB 4; Length 133;  
Best Local Similarity 83.3%; Pred. No. 71;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
| | | | |  
Db 14 RGRPEG 19

## RESULT 4

US-09-252-991A-22224  
; Sequence 22224, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22224  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22224

Query Match 91.2%; Score 31; DB 4; Length 217;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
| | | | |  
Db 198 RGRPQG 203

## RESULT 5

US-09-252-991A-19638  
; Sequence 19638, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19638  
; LENGTH: 291  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19638

Query Match 91.2%; Score 31; DB 4; Length 291;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
| | | | |  
Db 232 RGRPEG 237

## RESULT 6

US-09-489-039A-7834  
; Sequence 7834, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 7834  
; LENGTH: 423  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-7834

Query Match 91.2%; Score 31; DB 4; Length 423;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQG 6  
| | | | |  
Db 194 KGRPQG 199

## RESULT 7

US-09-238-303-8  
; Sequence 8, Application US/09238303B  
; Patent No. 6284253  
; GENERAL INFORMATION:  
; APPLICANT: Barr, Margaret C.  
; TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequence  
; FILE REFERENCE: 18617.0059  
; CURRENT APPLICATION NUMBER: US/09/238,303B  
; CURRENT FILING DATE: 1999-01-28  
; EARLIER APPLICATION NUMBER: US 60/072,927  
; EARLIER FILING DATE: 1998-01-29

NUMBER OF SEQ ID NOS: 17  
SEQ ID NO 8

LENGTH: 498

TYPE: PRT

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: protein encoded by the gag gene of a recombinant viral  
OTHER INFORMATION: clone constructed from the genomic DNA of a Pallas's cat feline  
OTHER INFORMATION: immunodeficiency virus  
US-09-238-303-8

Query Match 91.2%; Score 31; DB 3; Length 498;

Best Local Similarity 83.3%; Pred. No. 2.4e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6

Db 395 RGRPQG 400

RESULT 8

US-09-946-239-8

Sequence 8, Application US/09946239

Patent No. 6579527

GENERAL INFORMATION:

APPLICANT: Barr, Margaret C.

TITLE OF INVENTION: No. 6579527el Feline Immunodeficiency Virus Nucleotide and

FILE REFERENCE: 18617.0059

CURRENT APPLICATION NUMBER: US/09/946,239

PRIOR FILING DATE: 2001-09-04

PRIOR APPLICATION NUMBER: US 09/238,303, US 60/072,927

PRIOR FILING DATE: 1999-01-28, 1998-01-29

NUMBER OF SEQ ID NOS: 17

SEQ ID NO 8

LENGTH: 498

TYPE: PRT

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: protein encoded by the gag gene of a recombinant viral  
OTHER INFORMATION: clone constructed from the genomic DNA of a Pallas's cat feline  
OTHER INFORMATION: immunodeficiency virus  
US-09-946-239-8

Query Match 91.2%; Score 31; DB 4; Length 498;

Best Local Similarity 83.3%; Pred. No. 2.4e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6

Db 395 RGRPQG 400

RESULT 9

US-09-252-991A-19991

Sequence 19991, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 19991

LENGTH: 787

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-19991

Query Match

Best Local Similarity 91.2%; Score 31; DB 4; Length 787;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6

Db 299 RGRPQG 304

RESULT 10

US-09-252-991A-30397

Sequence 30397, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 30397

LENGTH: 894

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30397

Query Match

Best Local Similarity 91.2%; Score 31; DB 4; Length 894;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQG 6

Db 830 RGRPQG 835

RESULT 11

US-08-804-227C-12

Sequence 12, Application US/08804227C

Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: DeHoff, Bradley S.

APPLICANT: Kuhstoss, Stuart A.

APPLICANT: Rostock, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501

STREET: LILLY CORPORATE CENTER

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCI(DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804,227C

FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1891 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-227C-12

Query Match 91.2%; Score 31; DB 2; Length 1891;  
Best Local Similarity 83.3%; Pred. No. 8.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
Db 882 RGRPEG 887

RESULT 12  
US-08-804-198-6  
Sequence 6, Application US/08804198  
Patent No. 5945320  
GENERAL INFORMATION:  
APPLICANT: Burgett, Stanley G.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rac, Nagaraja R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Rostock, Paul R., Jr.  
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESS: PAUL R. CANTRELL 1138  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1891 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-198-6

Query Match 91.2%; Score 31; DB 2; Length 1891;  
Best Local Similarity 83.3%; Pred. No. 8.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
Db 882 RGRPEG 887

RESULT 13  
US-09-839-743-18  
Sequence 18, Application US/09839743

Patent No. 6472211  
GENERAL INFORMATION:  
APPLICANT: The Salk Institute for Biological Sciences  
APPLICANT: Lamb, Christopher  
APPLICANT: Doerner, Peter  
APPLICANT: Laible, Goetz  
TITLE OF INVENTION: No. 6472211el Transcription Enhancer Element and  
TITLE OF INVENTION: Transcription Factor and Methods of Use Thereof  
FILE REFERENCE: SALKINS.008DV3  
CURRENT APPLICATION NUMBER: US/09/839,743  
CURRENT FILING DATE: 2001-04-19  
PRIOR APPLICATION NUMBER: US 09/401,336  
PRIOR FILING DATE: 1999-09-21  
PRIOR APPLICATION NUMBER: US 09/189,344  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: US 08/669,721  
PRIOR FILING DATE: 1996-06-27  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 11  
TYPE: PRT  
ORGANISM: Nicotiana tabacum  
US-09-839-743-18

Query Match 88.2%; Score 30; DB 4; Length 11;  
Best Local Similarity 83.3%; Pred. No. 11;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
Db 6 RGRPKG 11

RESULT 14  
US-09-258-373-22  
Sequence 22, Application US/09258373  
Patent No. 6150110  
GENERAL INFORMATION:  
APPLICANT: Fletcher, Jonathan A.  
APPLICANT: Xiao, Sheng  
TITLE OF INVENTION: HMGI(Y)-LAMA4\* FUSION ONCOGENE,  
TITLE OF INVENTION: ONCOPROTEIN AND METHODS OF USE  
FILE REFERENCE: B0801/7135/ERP  
CURRENT APPLICATION NUMBER: US/09/258,373  
CURRENT FILING DATE: 1999-02-26  
EARLIER APPLICATION NUMBER: 60/076,401  
EARLIER FILING DATE: 1998-02-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 22  
LENGTH: 96  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-258-373-22

Query Match 88.2%; Score 30; DB 3; Length 96;  
Best Local Similarity 83.3%; Pred. No. 80;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOG 6  
Db 47 RGRPKG 52

RESULT 15  
US-09-341-444A-2  
Sequence 2, Application US/09341444A  
Patent No. 6440666  
GENERAL INFORMATION:  
APPLICANT: Groenen, Martinus Antonius Mathilda  
APPLICANT: Albers, Gerardus Antonius Arnoldus  
TITLE OF INVENTION: Selection For Dwarfism in Poultry

```
; FILE REFERENCE: 310-1009
; CURRENT APPLICATION NUMBER: US/09/341,444A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/NL98/00021
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: EP 97200070.7
; PRIOR FILING DATE: 1997-01-10
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Chicken
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (77)..(77)
; OTHER INFORMATION: X is unknown
US-09-341-444A-2

Query Match      88.2%; Score 30; DB 4; Length 109;
Best Local Similarity 83.3%; Pred. No. 90;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 RGRPQG 6
Db      49 RGRPKG 54

Search completed: April 6, 2004, 16:19:38
Job time : 8.80374 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 7.85047 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-3

Perfect score: 41

Sequence: 1 RGRPQGP 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_78.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	166	1	PIHUSC
2	41	100.0	166	2	salivary proline-r
3	41	100.0	171	2	salivary proline-r
4	37	90.2	188	2	proline-rich phosph
5	37	90.2	317	2	proline-rich phosph
6	37	90.2	740	1	proline-rich phosph
7	36	87.8	212	2	gag polypeptide
8	36	87.8	240	2	proline-rich phosph
9	36	87.8	240	2	proline-rich phosph
10	36	87.8	278	2	proline-rich phosph
11	35	85.4	415	2	proline-rich phosph
12	34	82.9	284	2	hypothetical protei
13	34	82.9	330	2	hypothetical protei
14	34	82.9	698	2	hypothetical protei
15	34	82.9	698	2	hypothetical protei
16	34	82.9	1137	1	ribonucleoside-dip
17	34	82.9	1159	2	probable potassium
18	34	82.9	1784	2	tuberculous sclerosi
19	34	82.9	2188	2	probable polyketid
20	33	80.5	76	2	basic proline-rich
21	33	80.5	117	2	basic proline-rich
22	33	80.5	128	2	basic proline-rich
23	33	80.5	141	2	hypothetical prote
24	33	80.5	141	2	hypothetical prote
25	33	80.5	147	2	hypothetical prote
26	33	80.5	164	2	proline-rich prote
27	33	80.5	170	2	proline-rich prote
28	33	80.5	172	2	proline-rich prote
29	33	80.5	188	2	basic proline-rich

## ALIGNMENTS

### RESULT 1

#### PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N:Alternate names: salivary acidic proline-rich protein PRH2

N:Contains: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotein

C:Species: Homo sapiens (man)

C>Date: 31-Mar-1991 #sequence revision 12-Apr-1996 #text change 08-Dec-2000

C:Accession: A25372; A19803; B57868; A92277; A92254; A94255; A91954; S02564; S02563; JP

C:Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A>Title: Differential RNA splicing and post-translational cleavages in the human saliva

A:Reference number: A92492; MUID:85289325; PMID:2993301

A:Accession: A25372

A:Molecule type: mRNA

A:Residues: 1-166 <MAE>

A:Cross-references: GB:X03020; NID:g190481; PIDN:AAA50183.1; PID:g190482

R:Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A>Title: Primary structure of the active tryptic fragments of human and monkey salivary

A:Reference number: A91757; MUID:81191179; PMID:7228490

A:Accession: A19803

A:Molecule type: protein

A:Residues: 17-46 <SCH>

R:Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A>Title: Structures of two HaeIII-type genes in the human salivary proline-rich protein

A:Reference number: A57868; MUID:86156106; PMID:3009472

A:Accession: B57868

A:Molecule type: DNA

A:Residues: 1-166 <KIM>

A:Cross-references: GB:M13058; NID:g190513; PIDN:AAA98808.1; PID:g190514

R:Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A>Title: The primary structure of a salivary calcium-binding proline-rich phosphoprotein

A:Reference number: A92277; MUID:80204369; PMID:7380845

A:Contents: protein C

A:Accession: A92277

A:Molecule type: protein

A:Residues: 17-19, 'N', 21-166 <WON>

A>Note: The amino-terminal 46 residues are involved with inhibiting hydroxyapatite form.

R:Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A>Title: The complete primary structure of a proline-rich phosphoprotein from human sal

A:Reference number: A92254; MUID:79173237; PMID:438215

A:Contents: protein A

A:Accession: A92254

A:Molecule type: protein

A:Residues: 17-19, 'N', 21-122 <WO2>

R:Schlesinger, D.H.; Hay, D.I.

In Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gross

A>Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A:Reference number: A94425

```

A:Accession: A94425
A:Molecule type: protein
A:Residues: 17-122 <SC2>
A>Note: The authors call this protein PRP-4
R:Isumura, S.; Saitoh, E.; Sanada, K.
J. Biochem. 87, 1071-1077, 1980
A>Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to the structure of a proline-rich protein
A:Reference number: A91954; MUID:80227634; PMID:7390979
A:Contents: peptide P-C
A:Accession: A91954
A:Molecule type: protein
A:Residues: 123-166 <ISE>
R:Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck
Biochem. J. 255, 15-21, 1988
A>Title: Structures of two HeaII-type genes in the human salivary proline-rich protein
A:Reference number: S02562; MUID:89061650; PMID:3196309
A:Accession: S02564
A:Molecule type: protein
A:Residues: 17-166 <HAY>
A:Accession: S02563
A:Molecule type: protein
A:Residues: 47-71 <HAZ>
R:Schlesinger, D.H.; Hay, D.I.
Int. J. Pept. Protein Res. 27, 373-379, 1986
A>Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of the primary structure and possible origin of the non-glycosylated basic proline-rich protein
A:Reference number: JP0106; MUID:86222916; PMID:3710693
A:Accession: JP0106
A:Molecule type: protein
A:Residues: 17-161, 'Q', 163-166 <SC3>
A:Experimental source: parotid gland
R:Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
Biochemistry 30, 3351-3356, 1991
A>Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure and possible origin of the non-glycosylated basic proline-rich protein
A:Reference number: A38355; MUID:91190884; PMID:1849422
A:Accession: G38355
A>Status: Preliminary
A:Molecule type: protein
A:Residues: 123-166 <KAU>
R:Robinson, R.; Kauffman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.
Biochem. J. 263, 497-503, 1989
A>Title: Primary structure and possible origin of the non-glycosylated basic proline-rich protein
A:Reference number: S06153; MUID:9008384; PMID:2688632
A:Accession: S06153
A:Molecule type: protein
A:Residues: 123-166 <ROB>
A:Comment: The proposed biological functions are a highly potent inhibitor of crystal growth
C:Genetics:
A:Gene: GDB:PRH2
A:Cross-references: GDB:119516; OMIM:168790
A:Map position: 12p13.2-12p13.2
A:Introns: 22/1, 34/1
C:Superfamily: proline-rich protein
C:Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva
F:1-16/Domain: signal sequence #status predicted <RG>
F:17-166/Product: protein C #status experimental <PRC>
F:17-122/Product: protein A #status experimental <PRA>
F:17-46/Region: apatitic mineral binding
F:47-71/Product: PRP-3 #status experimental <PRP3>
F:123-166/Product: peptide P-C #status experimental <PPC>
F:17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental
F:24,38/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 100.0%; Score 41; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7
DB 122 RGRPQGP 128

RESULT 3
A27307
proline-rich phosphoprotein (gene PRH1, Db allele) - human
N:Alternate names: salivary acidic proline-rich protein
C:Species: Homo sapiens (man)

```

---

```

salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human
C:Species: Homo sapiens (man)
C>Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text change 20-Aug-1999
C:Accession: B25372; A57868; S02562; G38355; S06153; B27307
R:Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.
J. Biol. Chem. 260, 11123-11130, 1985
A>Title: Differential RNA splicing and post-translational cleavages in the human salivary
A:Reference number: A9492; MUID:85289325; PMID:2993301
A:Accession: B25372
A:Molecule type: mRNA
A:Residues: 1-166 <MAE>
A:Cross-references: GB:K03203; NID:G190483; PIDN:AAA60184.1; PID:G190484
R:Kim, H.S.; Maeda, N.
J. Biol. Chem. 261, 6712-6718, 1986
A>Title: Structures of two HeaII-type genes in the human salivary proline-rich protein
A:Reference number: A57868; MUID:86196106; PMID:3009472
A:Accession: A57868
A:Molecule type: DNA
A:Residues: 1-166 <KIM>
A:Cross-references: GB:M13057; NID:G190511; PIDN:AAA98807.1; PID:G190512
R:Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck
Biochem. J. 255, 15-21, 1988
A>Title: The primary structures of six human salivary acidic proline-rich proteins (PRP
A:Reference number: S02562; MUID:89061650; PMID:3196309
A:Accession: S02562
A:Molecule type: protein
A:Residues: 47-71 <HAY>
R:Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
Biochemistry 30, 3351-3356, 1991
A>Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure and possible origin of the non-glycosylated basic proline-rich protein
A:Reference number: A38355; MUID:91190884; PMID:1849422
A:Accession: G38355
A:Molecule type: protein
A:Residues: 123-166 <KAU>
R:Robinson, R.; Kauffman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.
Biochem. J. 263, 497-503, 1989
A>Title: Primary structure and possible origin of the non-glycosylated basic proline-rich protein
A:Reference number: S06153; MUID:9008384; PMID:2688632
A:Accession: S06153
A:Molecule type: protein
A:Residues: 123-166 <ROB>
R:Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.
Am. J. Hum. Genet. 41, 1035-1047, 1987
A>Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein
A:Reference number: A27307; MUID:88074309; PMID:3687941
A:Contents: allele Pa
A:Accession: B27307
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 17-41, 'L', 43-118, 'C', 120-166 <AZE>
A:Cross-references: EMBL:K03203
C:Genetics:
A:Gene: GDB:PRH1
A:Cross-references: GDB:119515; OMIM:168730
A:Map position: 12p13.2-12p13.2
A:Introns: 22/1, 34/1
C:Superfamily: proline-rich protein
C:Keywords: phosphoprotein; saliva; tandem repeat

Query Match 100.0%; Score 41; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7
DB 122 RGRPQGP 128

RESULT 3
A27307
proline-rich phosphoprotein (gene PRH1, Db allele) - human
N:Alternate names: salivary acidic proline-rich protein
C:Species: Homo sapiens (man)

```

C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 29-Aug-1997  
 C:Accession: A27307  
 R:Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A:Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
 A:Reference number: A27307; MUID:88074309; PMID:3687941  
 A:Accession: A27307  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-171 <AZE>  
 A:Cross-references: EMBL:K03203  
 C:Genetics:  
 A:Gene: GDB:PRH1  
 A:Cross-references: GDB:119515; OMIM:168730  
 A:Map position: 12p13.2-12p13.2  
 C:Superfamily: proline-rich protein  
 C:Keywords: phosphoprotein

Query Match 100.0%; Score 41; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 2.4; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGP 7  
 |||||  
 Db 127 RGRPOGP 133

RESULT 4  
 D29149  
 proline-rich protein - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 20-Aug-1999  
 C:Accession: D29149  
 R:Clements, S.; Mehancho, H.; Carlson, D.M.  
 J. Biol. Chem. 260, 13471-13477, 1985  
 A:Title: Novel multigene families encoding highly repetitive peptide sequences. Sequence  
 A:Reference number: A92501; MUID:86033799; PMID:384080  
 A:Contents: Clone pUM24  
 A:Accession: D29149  
 A:Molecule type: mRNA  
 A:Residues: 1-188 <CLE>  
 A:Cross-references: GB:M19419; NID:g200542; PIDN:AAA40002.1; PID:g200543  
 C:Superfamily: proline-rich protein

Query Match 90.2%; Score 37; DB 2; Length 188;  
 Best Local Similarity 85.7%; Pred. No. 14; Mismatches 1; Indels 0; Gaps 0;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGP 7  
 |||||  
 Db 181 QGRPOGP 187

RESULT 5  
 A28996  
 proline-rich protein M14 precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 20-Aug-1999  
 C:Accession: A28996  
 R:Ann, D.K.; Smith, K.; Carlson, D.M.  
 J. Biol. Chem. 263, 10887-10893, 1988  
 A:Title: Molecular evolution of the mouse proline-rich protein multigene family. Inserti  
 A:Reference number: A28996; MUID:88273214; PMID:2839509  
 A:Accession: A28996  
 A:Molecule type: DNA  
 A:Residues: 1-317 <ANN>  
 A:Cross-references: GB:M23236; GB:J03891; NID:g200535; PIDN:AAA53048.1; PID:g567232  
 C:Genetics:  
 A:Introns: 22/1  
 C:Superfamily: proline-rich protein  
 C:Keywords: saliva  
 F:16-317/Product: proline-rich protein M14 #status predicted <SIG>

Query Match 90.2%; Score 37; DB 2; Length 317;  
 Best Local Similarity 85.7%; Pred. No. 22; Mismatches 1; Indels 0; Gaps 0;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGP 7  
 |||||  
 Db 310 QGRPOGP 316

RESULT 6  
 FOLJHD  
 gag polyprotein - squirrel monkey retrovirus SMRV-H  
 N:Contains: core protein p16; core protein p19; probable core protein p10; probable cor  
 C:Species: squirrel monkey retrovirus SMRV-H  
 C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999  
 C:Accession: A31827  
 R:Oda, T.; Ikeda, S.; Watanabe, S.; Matsushika, M.; Akiyama, K.; Mitsunobu, F.  
 Virology 167, 468-476, 1988  
 A:Title: Molecular cloning, complete nucleotide sequence, and gene structure of the pro  
 A:Reference number: A31827; MUID:89073750; PMID:3201749  
 A:Accession: A31827  
 A:Molecule type: DNA  
 A:Residues: 1-740 <ODA>  
 A:Cross-references: GB:M23385; NID:g332626; PIDN:AAA66451.1; PID:g807672  
 C:Genetics:  
 A:Gene: gag  
 C:Superfamily: AIDS-related virus gag polyprotein  
 C:Keywords: core protein; polyprotein  
 F:163/Product: core protein p19 #status predicted <CP9>  
 F:164-318/Product: core protein p16 #status predicted <CP6>  
 F:319-648/Product: core protein p35 #status predicted <CP5>  
 F:649-740/Product: core protein p10 #status predicted <CP1>

Query Match 90.2%; Score 37; DB 1; Length 740;  
 Best Local Similarity 85.7%; Pred. No. 47; Mismatches 1; Indels 0; Gaps 0;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGP 7  
 |||||  
 Db 681 RGRPOGP 687

RESULT 7  
 B36298  
 proline-rich protein PRB3S (cys) - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 29-Aug-1997  
 C:Accession: B36298  
 R:Azen, E.A.; Minaguchi, K.; Latreille, P.; Kim, H.S.  
 Am. J. Hum. Genet. 47, 686-697, 1990  
 A:Title: Alleles at the PRB3 locus coding for a disulfide-bonded human salivary proline  
 A:Reference number: A36298; MUID:91022705; PMID:2171329  
 A:Accession: B36298  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-212 <AZE>  
 C:Genetics:  
 A:Gene: GDB:PRB3  
 A:Cross-references: GDB:119513; OMIM:168840  
 A:Map position: 12p13.2-12p13.2  
 C:Superfamily: proline-rich protein

Query Match 97.8%; Score 36; DB 2; Length 212;  
 Best Local Similarity 100.0%; Pred. No. 23; Mismatches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RGRPOGP 7  
 |||||  
 Db 167 RGRPOGP 172

RESULT 8

B24264  
 C:Species: Mus musculus (house mouse)  
 C:Date: 09-Sep-1987 #sequence\_revision 09-Sep-1987 #text\_change 03-May-1996  
 C:Accession: B24264  
 R:Ann, D.K.; Carlson, D.M.  
 J. Biol. Chem. 260, 15863-15872, 1985  
 A:Title: The structure and organization of a proline-rich protein gene of a mouse multigene family  
 A:Reference number: A92508; MUID:86059475; PMID:2999141  
 A:Accession: B24264  
 A:Molecule type: DNA  
 A:Residues: 1-240 <ANN>  
 C:Superfamily: proline-rich protein

Query Match 87.8%; Score 36; DB 2; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7  
 |||||  
 Db 234 GRPQGP 239

RESULT 9  
 A24264  
 C:Species: Mus musculus (house mouse)  
 C:Date: 09-Sep-1987 #sequence\_revision 09-Sep-1987 #text\_change 03-May-1996  
 C:Accession: A24264  
 R:Ann, D.K.; Carlson, D.M.  
 J. Biol. Chem. 260, 15863-15872, 1985  
 A:Title: The structure and organization of a proline-rich protein gene of a mouse multigene family  
 A:Reference number: A92508; MUID:86059475; PMID:2999141  
 A:Accession: A24264  
 A:Molecule type: DNA  
 A:Residues: 1-240 <ANN>  
 C:Superfamily: proline-rich protein

Query Match 87.8%; Score 36; DB 2; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7  
 |||||  
 Db 234 GRPQGP 239

RESULT 10  
 B39066  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 20-Mar-1992 #sequence\_revision 10-Apr-1992 #text\_change 03-May-1996  
 C:Accession: B39066  
 R:Lin, H.H.; Ann, D.K.  
 Genomics 10, 102-113, 1991  
 A:Title: Molecular characterization of rat multigene family encoding proline-rich protein  
 A:Reference number: A39066; MUID:91257817; PMID:2045095  
 A:Accession: B39066  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-278 <LIN>  
 A:Cross-references: GB:M36414  
 C:Superfamily: proline-rich protein

Query Match 87.8%; Score 36; DB 2; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7  
 |||||  
 Db 267 GRPQGP 272

RESULT 11  
 T32490  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T32490  
 R:Ledwith, J.; Wohldmann, P.; Beck, C.  
 submitted to the EMBL Data Library, September 1997  
 A:Description: The sequence of C. elegans cosmid F09G2.  
 A:Reference number: Z21176  
 A:Accession: T32490  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-415 <LED>  
 A:Cross-references: EMBL:AF026215; PIDN:AA871137.1; GSPDB:GN00023; CESP:F09G2.9  
 A:Experimental source: strain Bristol N2; clone F09G2  
 C:Genetics:  
 A:Gene: CESP:F09G2.9  
 A:Map position: 5  
 A:Introns: 28/1; 57/3; 102/3; 300/3; 361/3

Query Match 85.4%; Score 35; DB 2; Length 415;  
 Best Local Similarity 85.7%; Pred. No. 64;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
 |||||  
 Db 231 RGRPQGP 237

RESULT 12  
 S74256  
 C:Species: Mus musculus (house mouse)  
 C:Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 02-Jul-1998  
 C:Accession: S74256  
 R:Kawakami, K.; Ohto, H.; Takizawa, T.; Saito, T.  
 FEBS Lett. 393, 259-263, 1996  
 A:Title: Identification and expression of six family genes in mouse retina.  
 A:Reference number: S74253; MUID:96409319; PMID:8814301  
 A:Accession: S74256  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-284 <KAW>  
 A:Cross-references: EMBL:D83147  
 A:Experimental source: retina; strain BALB/c  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regulat  
 F:208-264/Domain: homeobox homology <HOX>

Query Match 82.9%; Score 34; DB 2; Length 284;  
 Best Local Similarity 85.7%; Pred. No. 69;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
 |||||  
 Db 182 RGRPLGP 188

RESULT 13  
 S74255  
 C:Species: Mus musculus (house mouse)  
 C:Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 02-Jul-1998  
 C:Accession: S74255  
 R:Kawakami, K.; Ohto, H.; Takizawa, T.; Saito, T.  
 FEBS Lett. 393, 259-263, 1996  
 A:Title: Identification and expression of six family genes in mouse retina.  
 A:Reference number: S74253; MUID:96409319; PMID:8814301  
 A:Accession: S74255  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-330 <KAW>



A:Cross-references: EMBL:D83147  
 A:Experimental source: recina; strain BALB/c  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regulation  
 F:205-261/Domain: homeobox homology <HOX>

Query Match 82.9%; Score 34; DB 2; Length 330;  
 Best Local Similarity 85.7%; Pred. No. 79;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPQGP 7.  
 |||||  
 Db 179 RGRPLGP 185

RESULT 14  
 E85369  
 hypothetical protein AT4G31520 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: E85369  
 R:anonymous; The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
 Nature 402, 769-777, 1999  
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A:Reference number: A85001; MUID:20083488; PMID:10617198  
 A:Accession: E85369  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-698 <STO>  
 A:Cross-references: GB:NC\_001269; NID:g7270055; PIDN:CAB79870.1; GSPDB:GN00140  
 C:Genetics:  
 A:Gene: AT4G31520  
 A:Map position: 4

Query Match 82.9%; Score 34; DB 2; Length 698;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
 |||||  
 Db 413 RGRPQGP 419

RESULT 15  
 T10682  
 hypothetical protein F3L17.90 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Oct-1999  
 C:Accession: T10682  
 R:Sevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X  
 submitted to the Protein Sequence Database, June 1999  
 A:Reference number: Z16652  
 A:Accession: T10682  
 A:Molecule type: DNA  
 A:Residues: 1-698 <BEV>  
 A:Cross-references: EMBL:AL080283; GSPDB:GN00062; ATSP:F3L17.90  
 A:Experimental source: cultivar Columbia; BAC clone F3L17  
 C:Genetics:  
 A:Gene: ATSP:F3L17.90  
 A:Map position: 4  
 A:Introns: 51/3; 107/3; 152/2; 186/2; 226/3; 268/3; 300/3; 318/3; 343/3; 363/3; 402/3; 5

Query Match 82.9%; Score 34; DB 2; Length 698;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
 |||||  
 Db 413 RGRPQGP 419

Search completed: April 6, 2004, 16:16:50  
 Job time : 7.85047 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 4.51402 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-3

Perfect score: 41

Sequence: 1 RGRPQGP 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	41	100.0	166	1 PRP2_HUMAN	P02810 homo sapien
2	37	90.2	261	1 PRP2_MOUSE	P05142 mus musculus
3	37	90.2	740	1 GRG_SMRVH	P21411 squirrel mo
4	36	87.8	296	1 PMP3_MOUSE	P05143 mus musculus
5	35	85.4	124	1 CSS4_HUMAN	Q9bxq7 homo sapien
6	35	85.4	193	1 Y005_COREF	Q8ful3 corynebacte
7	34	82.9	148	1 RY225_HUMAN	Q9y225 homo sapien
8	34	82.9	199	1 RIMM_RH10	Q98ee0 rhizobium l
9	34	82.9	246	1 SIX6_CHICK	Q93307 gallus gall
10	34	82.9	246	1 SIX6_HUMAN	Q95475 homo sapien
11	34	82.9	246	1 SIX6_MOUSE	Q9qz28 mus musculus
12	34	82.9	272	1 SIX3_ORYLA	O73916 oryzias lat
13	34	82.9	314	1 SIX3_CHICK	O42406 gallus gall
14	34	82.9	331	1 KDD2_STRAW	O82bx4 streptomyce
15	34	82.9	332	1 SIX3_HUMAN	Q95343 homo sapien
16	34	82.9	333	1 SIX3_MOUSE	Q62233 mus musculus
17	34	82.9	374	1 YF01_RHME	Q92q49 rhizobium m
18	34	82.9	773	1 AD11_MOUSE	Q9riv4 mus musculus
19	34	82.9	1137	1 RIR1_HSV11	P08543 herpes simp
20	34	82.9	1158	1 KCH2_CANPA	Q9tsz3 canis fami
21	34	82.9	1159	1 KCH2_HUMAN	Q12809 homo sapien
22	34	82.9	1161	1 KCH2_RABIT	O8wny2 oryctolagus
23	34	82.9	1807	1 TSC2_HUMAN	P49815 homo sapien
24	33	80.5	61	1 PRP5_HUMAN	P02811 homo sapien
25	33	80.5	96	1 PRP2_RAT	P04281 homo sapien
26	33	80.5	172	1 PRP2_MOUSE	P10164 rattus norv
27	33	80.5	174	1 PRP3_MOUSE	P81489 homo sapien
28	33	80.5	206	1 PRP3_RAT	P04474 rattus norv
29	33	80.5	234	1 PRP4_HUMAN	P10161 homo sapien
30	33	80.5	247	1 PRP4_MOUSE	P10163 homo sapien
31	33	80.5	251	1 PRP2_HUMAN	P02812 homo sapien
32	33	80.5	266	1 SC02_HUMAN	O43819 homo sapien
33	33	80.5	271	1 TRMB_CAUCR	P58088 caulobacter

RESULT 1					
ID	PRPC_HUMAN	STANDARD;	PRT;	166	AA.
AC	P02810;				
DT	21-JUL-1986	(Rel. 01, Created)			
DT	01-MAR-1989	(Rel. 10, Last sequence update)			
DT	15-MAR-2004	(Rel. 43, Last annotation update)			
DE	Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains:				
DE	Peptide P-C].				
DE	PRH1 AND PRH2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).				
RX	MEDLINE=86196106; PubMed=3009472;				
RA	Kim H.-S., Maeda N.;				
RT	"Structures of two Haellii-type genes in the human salivary				
RT	proline-rich protein multigene family.";				
RL	J. Biol. Chem. 261:6712-6718(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).				
RX	MEDLINE=85289325; PubMed=2993301;				
RA	Maeda N., Kim H.-S., Azen E.A., Smithies O.;				
RT	"Differential RNA splicing and post-translational cleavages in the				
RT	human salivary proline-rich protein gene system.";				
RL	J. Biol. Chem. 260:11123-11130(1985).				
RN	[3]				
RP	SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).				
RX	MEDLINE=89061650; PubMed=3196309;				
RA	Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K.,				
RA	Madapallamattam G., Schluckebier S.K.;				
RT	"The primary structures of six human salivary acidic proline-rich				
RT	proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";				
RL	Biochem. J. 255:15-21(1988).				
RN	[4]				
RP	SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).				
RX	MEDLINE=88074309; PubMed=3687941;				
RA	Azen E.A., Kim H.-S., Goodman P., Flynn S., Maeda N.;				
RT	"Alleles at the PRH1 locus coding for the human salivary-acidic				
RT	proline-rich proteins Pa, Db, and PIF.";				
RL	Am. J. Hum. Genet. 41:1035-1047(1987).				
RN	[5]				
RP	SEQUENCE OF 17-166 (PRP-2).				
RX	MEDLINE=86222916; PubMed=3710693;				
RA	Schlesinger D.H., Hay D.I.;				
RT	"Complete covalent structure of a proline-rich phosphoprotein, PRP-2,				
RT	an inhibitor of calcium phosphate crystal growth from human parotid				
RT	saliva.";				
RL	Int. J. Pept. Protein Res. 27:373-379(1986).				
RN	[6]				
RP	SEQUENCE OF 17-166 (PROTEIN C).				
RX	MEDLINE=80204368; PubMed=7380845;				
RA	Wong R.S.C., Bennick A.;				

P10162 homo sapien  
P04280 mus musculus  
Q96xw0 mus musculus  
Q9yft9 aeropyrum p  
Q9b9y0 homo sapien  
P03168 ground squi  
P09403 thermus the  
Q99952 homo sapien  
Q90611 gallus gall  
P77894 mycobacteri  
Q9nhv9 drosophila  
P33479 pseudorabie

#### ALIGNMENTS

RT "The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A." J. Biol. Chem. 255:5943-5948(1980).

RL [7]

RN SEQUENCE OF 17-46 (PROTEIN C).

RP MEDLINE=81191179; PubMed=7228490;

RX Schlesinger D.H., Hay D.I.;

RA "Primary structure of the active tryptic fragments of human and monkey salivary anionic proline-rich proteins." Int. J. Pept. Protein Res. 17:34-41(1981).

RL [8]

RN SEQUENCE OF 17-122 (PROTEIN A).

RP MEDLINE=79173237; PubMed=438215;

RX Wong R.S.C., Hofmann T., Bennick A.;

RA "The complete primary structure of a proline-rich phosphoprotein from human saliva." J. Biol. Chem. 254:4800-4808(1979).

RL [9]

RN SEQUENCE OF 17-122 (PROTEIN A).

RP Schlesinger D.H., Hay D.I.;

RA "Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent inhibitor of calcium phosphate precipitation in human parotid saliva." (In) Gross E., Meienhofer J. (eds.);

RL Peptides: structure and biological function (Proceedings of the 6th American peptide symposium), pp.133-136, Pierce Chemical Co., Rockford IL. (1979).

RL [10]

RN SEQUENCE OF 123-166 (PEPTIDE P-C).

RP MEDLINE=80227634; PubMed=7390979;

RX Isemura S., Saitoh E., Sanada K.;

RA "The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to a salivary proline-rich phosphoprotein, protein C." J. Biochem. 97:1071-1077(1980).

RL [11]

RN VARIANT PRH2-3 LYS-163.

RP Azen E.A.;

RA "A frequent mutation in the acidic proline-rich protein gene, PRH2, causing a Q147K change closely adjacent to the bacterial binding domain of the cognate salivary PRP (Pri') in Afro-Americans." Hum. Mutat. 12:72-72(1998).

CC -!- FUNCTION: PRP's act as highly potent inhibitors of crystal growth of calcium phosphates. They provide a protective and reparative environment for dental enamel which is important for the integrity of the teeth.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIF-S yield PRP-4, PRP-3 (protein A), and PIF-F, respectively.

CC -!- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele; allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is also known as PR1.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; K03202; AAA60183.1; -

DR EMBL; K03203; AAA60184.1; -

DR EMBL; M13057; AAA98807.1; -

DR EMBL; M13058; AAA98808.1; -

DR Genew; HGNC:9366; PRH1.

DR Genew; HGNC:9367; PRH2.

DR MIM; 168730; -

DR MIM; 168790; -

DR MIM; 168710; -

DR CO; GO:0005615; Cxextracellular space; TAS.

DR Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism;

KW Pyrrrolidone carboxylic acid.

FT SIGNAL 1 16 SALIVARY ACIDIC PROLINE-RICH

FT CHAIN 17 166 PHOSPHOPROTEIN 1/2.

FT CHAIN 17 122 SALIVARY ACIDIC PROLINE-RICH

FT CHAIN 123 166 PHOSPHOPROTEIN 3/4.

FT DOMAIN 17 46 PEPTIDE P-C.

FT MOD\_RES 17 17 INHIBIT HYDROXYAPATITE FORMATION, BIND

FT MOD\_RES 24 24 TO HYDROXYAPATITE AND CALCIUM.

FT MOD\_RES 38 38 PYRROLIDONE CARBOXYLIC ACID.

FT VARIANT 20 20 PHOSPHORYLATION.

FT VARIANT 66 66 D -> N (in allele PRH1-4).

FT VARIANT 163 163 D -> N (in allele PRH2-1).

FT VARIANT 41 41 Q -> K (in allele PRH2-3).

FT CONFLICT 41 41 F -> P (IN REF. 10).

FT SEQUENCE 166 AA; 1701.7 MW; A7DF62BF94E3C3EF CRC64;

SQ

Query Match 100.0%; Score 41; DB 1; Length 166;

Best Local Similarity 100.0%; Pred. No. 0.93;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7

DB 122 RGRPQGP 128

RESULT 2

PRP2\_MOUSE STANDARD; PRT; 261 AA.

ID AC P05142;

DT 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Proline-rich protein MP-2 precursor.

DE PRH1 OR PRP.

GN Mus musculus (Mouse).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP MEDLINE=86059475; PubMed=2999141;

RA Ann D.K., Carlson D.M.;

RT "The structure and organization of a proline-rich protein gene of a mouse multigene family." J. Biol. Chem. 260:15863-15872(1985).

RL CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M12099; AAA40004.1; -

DR MGD; MGI:97773; Prhl.

KW Repeat; Signal.

FT SIGNAL 1 15 POTENTIAL.

FT CHAIN 16 261 PROLINE-RICH PROTEIN MP-2.

FT SEQUENCE 261 AA; 26034 MW; 36E13BA7387F47D4 CRC64;

SQ

Query Match 90.2%; Score 37; DB 1; Length 261;

Best Local Similarity 85.7%; Pred. No. 8;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7

DB 254 RGRPQGP 260

```

RESULT 3
GAG_SMRVH
ID_GAG_SMRVH STANDARD; PRT; 740 AA.
AC F21411;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE GAG polypeptide (Contains: Core protein P19, Core protein P16;
DE Probable core protein P35; Probable core protein P10).
CN GAG.
OS Squirrel monkey retrovirus (SMRV-H) (SMRV-HLB).
OC Viruses; Retroviral viruses; Retroviridae; Betaretrovirus.
OX NCBI_TaxID=11856;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89073750; PubMed=3201749;
RA Oda T., Ikeda S., Watanabe S., Hatsushika M., Akiyama K.,
RA Mitsunobu F.;
RT "Molecular cloning, complete nucleotide sequence, and gene structure
RT of the provirus genome of a retrovirus produced in a human
RT lymphoblastoid cell line.";
RL Virology 167:468-476(1988).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M23385; AAA66451.1; -
CC PIR; A31827; FOLJHD.
DR
DR InterPro; IPR003322; Gag p10.
DR InterPro; IPR000721; Gag p24.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF02337; Gag p10; 1.
DR Pfam; PF00607; Gag p24; 1.
DR Pfam; PF00098; zf_CCHC; 1.
DR SMART; SM00343; Znf_CCHC; 1.
DR Core protein; Polyprotein.
KW CHAIN
FT CHAIN 1 163 CORE PROTEIN P19.
FT CHAIN 164 318 CORE PROTEIN P16.
FT CHAIN 319 648 PROBABLE CORE PROTEIN P35.
FT CHAIN 649 740 PROBABLE CORE PROTEIN P10.
SQ SEQUENCE 740 AA; 80543 MW; CC2503C2661221F5 CRC64;

Query Match 90.2%; Score 37; DB 1; Length 740;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPGQP 7
Db 681 RGPGQP 687

RESULT 4
FMP3_MOUSE
ID_FMP3_MOUSE STANDARD; PRT; 296 AA.
AC P05143;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Proline-rich protein MP-3 (Fragment).
CN PRP1 OR PRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86059475; PubMed=2999141;
RA Ann D.K., Carlson D.M.;
RT "The structure and organization of a proline-rich protein gene of a
RT mouse multigene family.";
RL J. Biol. Chem. 260:15863-15872(1985).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M12100; AAA40005.1; -
CC MGD; MGI:97773; Prhl.
DR
DR NON TER 1
SQ SEQUENCE 296 AA; 29521 MW; 7F146824E8AF3269 CRC64;

Query Match 87.8%; Score 36; DB 1; Length 296;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPGQP 7
Db 290 GRPGQP 295

RESULT 5
CES4_HUMAN
ID_CES4_HUMAN STANDARD; PRT; 124 AA.
AC Q9PXQ7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cat eye syndrome critical region protein 4 (Fragment).
CN CECR4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21275466; PubMed=11381032;
RA Fouts T.K., Brinkman-Mills P., Banting G.S., Maier S.A., Riaz M.A.,
RA Bridgland L.J., Hu S., Birren B., Minoshima S., Shimizu N., Pan H.,
RA Nguyen T., Fang F., Fu Y., Ray L., Wu H., Shaul S., Phan S., Yao Z.,
RA Chen F., Huan A., Hu P., Wang Q., Loh P., Qi S., Roe B.A.,
RA McDermid H.E.;
RT "Analysis of the cat eye syndrome critical region in humans and the
RT region of conserved synteny in mice: a search for candidate genes at
RT or near the human chromosome 22 pericentromere.";
RL Genome Res. 11:1053-1070(2001).
CC -1- TISSUE SPECIFICITY: Adult heart and skeletal muscle. Widely
CC expressed in fetal tissues.
CC -1- MISCELLANEOUS: Candidate gene for the Cat Eye Syndrome (CES), a
CC developmental disorder associated with the duplication of a 2 Mb
CC region of 22q11.2. Duplication usually takes in the form of a
CC supernumerary bisected isodicentric chromosome, resulting in
CC four copies of the region (represents an inv dup(22)(q11)). CES is
CC characterized clinically by the combination of coloboma of the
CC iris and anal atresia with fistula, downslanting palpebral
CC fissures, preauricular tags and/or pits, frequent occurrence of
CC heart and renal malformations, and normal or near-normal mental
CC development.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

1C entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

1R EMBL; AF307448; AAK30048.1; -  
1R Genew; HGN1842; CEK4.  
1T NON TER 1  
1Q SEQUENCE 124 AA; 12838 MW; 938E00386308EC7A CRC64;

Query Match 85.4%; Score 35; DB 1; Length 124;  
Best Local Similarity 85.7%; Pred. No. 8.7;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1RGRPOGP 7  
|||||  
58 RGRPOGP 64

## RESULT 6

1D Y005 COREP STANDARD; PRT; 193 AA.  
1C Q8FUL3;  
1T 10-OCT-2003 (Rel. 42, Created)  
1T 10-OCT-2003 (Rel. 42, Last sequence update)  
1T 10-OCT-2003 (Rel. 42, Last annotation update)  
1E Hypothetical UPF0232 protein CE0005.  
1N CE0005.  
1S Corynebacterium efficiens.  
1S Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
1C Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
1X NCBI\_TaxID=152794;  
1N [1]

SEQUENCE FROM N.A.  
STRAIN=IS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
MEDLINE=22723752; PubMed=12840036;

1A Nishio Y., Nakamura Y., Kawarabayashi Y., Usuda Y., Kimura E.,  
1A Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,  
1A Gojobori T.;  
1T "Comparative complete genome sequence analysis of the amino acid  
1T replacements responsible for the thermostability of Corynebacterium  
1T efficiens.";

1R Genome Res. 13:1572-1579(2003).  
1C -!- SIMILARITY: Belongs to the UPF0232 family.  
1C This SWISS-PROT entry is copyright. It is produced through a collaboration  
1C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
1C the European Bioinformatics Institute. There are no restrictions on its  
1C use by non-profit institutions as long as its content is in no way  
1C modified and this statement is not removed. Usage by and for commercial  
1C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
1C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

1R EMBL; AP005214; BAC16815.1; -  
1R HAMAP; MF 00630; -; 1  
1R InterPro; IPR007922; DUF721.  
1R Pfam; PF05258; DUF721; 1.  
1X Hypothetical protein; Complete proteome.  
1Q SEQUENCE 193 AA; 21708 MW; A52D39B19990C4E39 CRC64;

Query Match 85.4%; Score 35; DB 1; Length 193;  
Best Local Similarity 85.7%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1RGRPOGP 7  
|||||  
62 RGRPOGP 68

## RESULT 7

1D Y005 COREP STANDARD; PRT; 148 AA.  
1C Q9Y225; Q9Y225; Q9Y225;  
1T 28-FEB-2003 (Rel. 41, Created)  
1T 28-FEB-2003 (Rel. 41, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)  
RING finger protein 24.  
GN RNF24.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
SEQUENCE FROM N.A.  
MEDLINE=21638749; PubMed=11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Beare D.M., Beare D.M.,  
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Leharvalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConnachie L.J., McLeay K., McMurray A.A.,  
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,  
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
RA Szuca C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilmberg L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
Rogers J.;  
RA "The DNA sequence and comparative analysis of human chromosome 20.";  
RT Nature 414:865-871(2001).  
RN [2]  
SEQUENCE FROM N.A.  
TISSUE=Blood, and Kidney;  
MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,  
RA Altschul S.F., Zeeberg B., Suetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Barra N.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
SEQUENCE OF 45-148 FROM N.A.  
RA Basi M.T., Banfi S., Riboni M., Ballabio A., Borsani G.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL036778; CAB46627.1; -
DR EMBL; AL031670; CAB43182.1; -
DR EMBL; BC000213; AAH00213.1; -
DR EMBL; BC039584; AAH39584.1; -
DR EMBL; AL079313; CAB45279.1; -
DR HSSP; P28990; 1CHC
DR Genew; HGNC:13779; RNF24.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Zinc-finger. 78 119 RING-TYPE.
FT ZN_RING 45 45
FT CONFLICT 45 45 Y -> S (IN REF. 3).
SQ SEQUENCE 148 AA; 17209 MW; 66C240C3A5991EAS CRC64;

Query Match 82.9%; Score 34; DB 1; Length 148;
Best Local Similarity 85.7%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPQGP 7
DB 134 RGPQGP 140

RESULT 8
RIMM_RHILO STANDARD; PRT; 199 AA.
AC Q98EE0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable 16S rRNA processing protein RIMM.
GN RIMM OR MLH4288.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
CC -!- FUNCTION: Essential for efficient processing of 16S rRNA. Probably
CC part of the 30S subunit prior to or during the final step in the
CC processing of 16S free 30S ribosomal subunits, it could be some
CC accessory protein needed for efficient assembly of the 30S
CC subunit. It is needed in a step prior to rbfA during the
CC maturation of 16S rRNA. It has affinity for free ribosomal 30S
CC subunits but not for 70S ribosomes (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the RIMM family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF003003; BAB50980.1; -
DR HAMAP; MF_00014; -; 1.

DR InterPro; IPR007903; PRC_barrel.
DR InterPro; IPR002676; RIMM.
DR Pfam; PF05239; PRC; 1.
DR Pfam; PF01782; RIMM; 1.
DR RNA processing; Complete proteome.
SQ SEQUENCE 199 AA; 21156 MW; 321328B2BEA558CD CRC64;

Query Match 82.9%; Score 34; DB 1; Length 199;
Best Local Similarity 71.4%; Pred. No. 21;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7
DB 186 KGRPRGP 192

RESULT 9
SIX6_CHICK STANDARD; PRT; 246 AA.
AC Q93307;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeobox protein Six6 (Sixe oculis homeobox homolog 6) (Optic homeobox
DE 2) (Six9 protein).
GN SIX6 OR OPTX2 OR Six9.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC TISSUE=Embryonic retina;
RX MEDLINE=98393698; PubMed=9724757;
RA Toy J., Yang J.-M., Leppert G.S., Sundin O.H.;
RA "The optx2 homeobox gene is expressed in early precursors of the eye
RT and activates retina-specific genes."
RL Proc. Natl. Acad. Sci. U.S.A. 95:10643-10648(1998).
RN [2]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=99310672; PubMed=10381575;
RA Lopez-Rios J., Gallardo E., Rodriguez de Cordoba S., Bovolenta P.;
RA "Six9 (Optx2), a new member of the Six gene family of transcription
RT factors, is expressed at early stages of vertebrate ocular and
RT pituitary development."
RL Mech. Dev. 83:155-159(1999).
CC -!- FUNCTION: May be involved in eye development.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: In the developing embryo, expressed in the
CC anterior head-fold, the anterior neural plate and optic vesicle.
CC At later stages expression is maintained in the eye, while brain
CC expression becomes limited. Not expressed in the lens placode.
CC -!- DEVELOPMENTAL STAGE: Expression is first detected in the
CC prechordal mesoderm of stage 4 gastrulas.
CC -!- SIMILARITY: Belongs to the SIX/Sine oculis homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF050131; AAC33851.1; -
DR EMBL; AJ011786; CAA09774.1; -
DR HSSP; P41778; 1DU6.
DR GO; GO:0005634; C:nucleus; ISS.
DR GO; GO:0003677; F:DNA binding; ISS.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000047; HTH_lambrepreser.
```

R R R R R R W T Q  
y y b RE II ID AC DT DT DE DE EN DS DC DX RN RF RC RP RT RT RI RN RE RA RA RA RA RA RT RT RI RN RE RA RA RA RA RA RT RT RI RN RE RA RA RA RA RA CC CC CC CC CC

```

P SEQUENCE FROM N.A. TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
C STRAIN=BALB/c; TISSUE=Embryonic head;
X MEDLINE=99310672; PubMed=10381575;
A Lopez-Rios J., Gallardo E., Rodriguez de Cordoba S., Bovolenta P.;
T "Six9 (Optx2)", a new member of the Six gene family of transcription
T factors, is expressed at early stages of vertebrate ocular and
T pituitary development."
L Mech. Dev. 83:155-159(1999).
C -1- FUNCTION: May be involved in eye development.
C -1- TISSUE SPECIFICITY: Nuclear (By similarity).
C the ventral optic stalk, optic chiasma, the neural retina and the
C primordial tissues that give rise to the pituitary/hypothalamus
C axis. Not expressed in the lens placode.
C -1- DEVELOPMENTAL STAGE: Expression is first detected in the embryo at
C E8.
C -1- SIMILARITY: Belongs to the SIX/Sine oculis homeobox family.
C -1- SIMILARITY: Contains 1 homeobox domain.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
R EMBL; AF050130; AAC33850.1; -
R EMBL; AF135267; AAD48911.1; -
R EMBL; AJ011787; CAA09775.1; -
R EMBL; AK017544; -; NOT_ANNOTATED_CDS.
R HSP; P40424; I372.
R TRANSPAC; T03272; -
R MGD; MGI.1341840; Six6.
R GO; GO:0005634; C:nucleus; ISS.
R GO; GO:0003677; F:DNA binding; ISS.
R GO; GO:0005515; F:protein binding; IPI.
R InterPro; IPR001356; Homeobox.
R InterPro; IPR000047; HTH_lambrepres.
R InterPro; IPR007105; SIX.
R InterPro; IPR007106; SIX_SINE_homeo.
R Pfam; PF00046; homeobox; 1.
R PRINTS; PR00031; HTHREPRESSR.
R PRODOM; PD000010; Homeobox; 1.
R SMART; SM00389; HOX; 1.
R PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
R PROSITE; PS0071; HOMEBOX_2; 1.
W Developmental protein; Homeobox; DNA-binding; Nuclear protein.
T DNA BIND 126 186 HOMEBOX
T CONFLICT 141 141 H -> N (IN REF. 3).
T CONFLICT 220 220 S -> T (IN REF. 3).
T SEQUENCE 246 AA; 27741 MW; F1332D5E617B2CF1 CRC64;
Q
Query Match 82.9%; Score 34; DB 1; Length 246;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 RGRPQGP 7
b 103 RGRPLGP 109

RESULT 12
SIX3_ORYLA STANDARD; PRT; 272 AA.
AC 073916;
YT 15-DEC-1998 (Rel. 37, Created)
YT 15-DEC-1998 (Rel. 37, Last sequence update)
YT 10-OCT-2003 (Rel. 42, Last annotation update)
ZE Homeobox protein SIX3 (Sine oculis homeobox homolog 3).
EN SIX3.
XS Oryzias latipes (Medaka fish) (Japanese ricefish).
XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Query Match 82.9%; Score 34; DB 1; Length 246;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 RGRPQGP 7
b 103 RGRPLGP 109

RESULT 13
SIX3_CHICK STANDARD; PRT; 314 AA.
AC 042406;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeobox protein SIX3 (Sine oculis homeobox homolog 3) (CSIX3).
GN SIX3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=98168856; PubMed=9510037;
RA Bovolenta P., Mallamaci A., Puelles L., Boncinelli E.;
RT "Expression pattern of cSix3, a member of the Six/sine oculis family
RT of transcription factors."
RL Mech. Dev. 70:201-203(1998).

```



!!- FUNCTION: May be involved in visual system development.  
!!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
!!- SIMILARITY: Belongs to the SIX/Sine oculis homeobox family.  
!!- SIMILARITY: Contains 1 homeobox domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: Y15106; CAA75380.1; -.  
HSP: P41778; IDU6.  
TRANSFAC: T03482; -.  
InterPro: IPR001356; Homeobox.  
InterPro: IPR000047; HTH\_lambrepresr.  
InterPro: IPR007105; SIX.  
InterPro: IPR007106; SIX\_SINE\_homeo.  
Pfam: PF00046; homeobox\_1.  
PRINTS: PR00031; HTHREPRESSR.  
PRODOM: PD00010; Homeobox; 1.  
SMART: SM00389; HOX; 1.  
PROSITE: PS00027; HOMEBOX\_1; FALSE\_NEG.  
PROSITE: PS50071; HOMEBOX\_2; 1.  
Developmental protein; Homeobox; DNA-binding; Nuclear protein.  
DOMAIN 44 50 POLY-GLY.  
DNA\_BIND 188 247 HOMEBOX.  
DOMAIN 245 248 POLY-ALA.  
SEQUENCE 314 AA; 34677 MW; D9A04530185BA75F CRC64;  
-----  
Query Match 82.9%; Score 34; DB 1; Length 314;  
Best Local Similarity 85.7%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
-----  
Y 1 RGRPQGP 7  
| | | | |  
b 163 RGRPLGP 169  
-----  
RESULT 14  
JD2 STRAW STANDARD; PRT; 331 AA.  
AC Q82B4;  
IT 10-OCT-2003 (Rel. 42, Created)  
VT 10-OCT-2003 (Rel. 42, Last sequence update)  
VT 10-OCT-2003 (Rel. 42, Last annotation update)  
JE Probable 5-dehydro-4-deoxyglucuronate dehydratase 2 (EC 4.2.1.41) (5-  
DE keto-4-deoxy-glucuronate dehydratase 2) (KGDH 2).  
SN SAV5580.  
DS Streptomyces avermitilis.  
DC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
DC Streptomycineae; Streptomycetaceae; Streptomyces.  
DX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,  
RA "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
RT metabolites."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.,  
RA "Complete genome sequence and comparative analysis of the industrial  
RT

microorganism Streptomyces avermitilis."  
RL Nat. Biotechnol. 21:526-531(2003).  
CC !!- CATALYTIC ACTIVITY: 5-dehydro-4-deoxy-D-glucuronate = 2,5-  
CC dioxopentanoate + H(2)O + CO(2).  
CC !!- PATHWAY: Glucuronate catabolism; second step.  
CC !!- SIMILARITY: Belongs to the DHDPS family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: AP005043; BAC73292.1; -.  
HAMAP: MF\_00694; -; 1.  
InterPro: IPR002220; DHDPS.  
Pfam: PF00701; DHDPS; 1.  
PRODOM: PD01859; DHDPS; 1.  
KW Lyase; Complete proteome.  
SQ SEQUENCE 331 AA; 34707 MW; 8E1C7F345539D1C4 CRC64;  
-----  
Query Match 82.9%; Score 34; DB 1; Length 331;  
Best Local Similarity 85.7%; Pred. No. 36;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
-----  
QY 1 RGRPQGP 7  
| | | | |  
DB 294 RGRPVGP 300  
-----  
RESULT 15  
SIX3\_HUMAN STANDARD; PRT; 332 AA.  
ID SIX3\_HUMAN  
AC Q95343;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Homeobox protein SIX3 (Sine oculis homeobox homolog 3).  
GN SIX3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99107815; PubMed=9889003;  
RA Granadino B., Gallardo M.E., Lopez-Rios J., Sanz R., Ramos C.,  
RA Ayuso C., Bovolenta P., Rodriguez de Cordoba S.,  
PT "Genomic cloning, structure, expression pattern, and chromosomal  
RT location of the human SIX3 gene."  
RL Genomics 55:100-105(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99377859; PubMed=10454822;  
RA Leppert G.S., Yang J.-M., Sundin O.H.,  
PT "Sequence and location of SIX3, a homeobox gene expressed in the human  
RT eye."  
RL Ophthalmic Genet. 20:1-15(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retina;  
RA Clark B.J., Hanson I.M., Brown A.G., Ferrier R.K., Prosser J.,  
RA van Heyningen V.,  
RT "SIX3, a member of the Sine oculis/Six family of transcription  
RT factors, is expressed in the developing and adult human eye."  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP VARIANTS HPE2 VAL-226; ALA-250 AND PRO-257.  
RX MEDLINE=99295940; PubMed=10369266;  
RA Wallis D.E., Roessler E., Hehr U., Nanni L., Wiltshire T.,  
RA Richieri-Costa A., Gillesen-Kaesbach G., Zackai E.H., Rommens J.,

RA Muenke M.;  
RT "Mutations in the homeodomain of the human SIX3 gene cause  
RL holoprosencephaly.";  
RL Nat. Genet. 22:196-198(1999).  
CC -!- FUNCTION: May be involved in visual system development.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- DISEASE: Defects in SIX3 are the cause of holoprosencephaly type 2  
CC (HPE2) [MIM:157170]. HPE2 is a common, severe malformation of the  
CC brain that involves separation of the central nervous system into  
CC left and right halves.  
CC -!- SIMILARITY: Belongs to the SIX/Sine oculis homeobox family.  
CC -!- SIMILARITY: Contains 1 homeobox domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF092047; AAD11939.1; -;  
DR EMBL; AF049339; AAD15753.1; -;  
DR EMBL; AF083891; AAD51091.1; -;  
DR EMBL; AJ013611; CAB42539.1; -;  
DR HSSP; P41778; 1DU6.  
DR TRANSFAC; T03282; -;  
DR Genew; HGNC:10889; SIX3.  
DR MIM; 603714; -;  
DR MIM; 157170; -;  
DR GO; GO:0007420; P:brain development; TAS.  
DR GO; GO:0007601; P:vision; TAS.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR000047; HTH\_lambdarepressr.  
DR InterPro; IPR007105; SIX.  
DR InterPro; IPR007106; SIX\_SINE\_homeo.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00031; HTHREPRESSR.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; FALSE\_NEG.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Developmental protein; Homeobox; DNA-binding; Nuclear protein;  
KW Disease mutation; Holoprosencephaly.  
FT DOMAIN 33 69 POLY-GLY.  
FT DNA\_BIND 206 265 HOMEBOX.  
FT DOMAIN 263 266 POLY-ALA.  
FT VARIANT 226 226 L->V (in HPE2).  
FT VARIANT 250 250 /FTID=VAR\_003771.  
FT VARIANT 257 257 V->A (in HPE2).  
FT VARIANT 257 257 /FTID=VAR\_003772.  
FT VARIANT 257 257 R->P (in HPE2).  
FT VARIANT 257 257 /FTID=VAR\_003773.  
SQ SEQUENCE 332 AA; 35486 MW; 21EA07F6A2DD978F CRC64;

Query Match 82.9%; Score 34; DB 1; Length 332;  
Best Local Similarity 85.7%; Pred.No. 36;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 RGRPQGP 7  
Db 181 RGRPLGP 187

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 25.3178 Seconds

(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-3

Perfect score: 41

Sequence: 1 RGRPOGP 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

'SPTREMBL\_25':

- 1: sp\_archea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	256	16 Q7UYG7	Q7UYG7 rhodopirell
2	37	90.2	188	11 Q62106	Q62106 mus musculus
3	37	90.2	317	11 Q62103	Q62103 mus musculus
4	37	90.2	393	15 Q88189	Q88189 squirrel mo
5	36	87.8	249	2 Q812K2	Q812K2 moorella th
6	36	87.8	274	11 Q04154	Q04154 rattus norv
7	36	87.8	439	10 Q857B4	Q857B4 oryza sativ
8	36	87.8	439	10 Q7XFL9	Q7XFL9 oryza sativ
9	36	87.8	579	2 Q8GFF2	Q8GFF2 streptomyce
10	36	87.8	603	16 Q9RL17	Q9RL17 streptomyce
11	36	87.8	803	16 Q81ZW2	Q81ZW2 streptomyce
12	36	87.8	917	10 Q8FT90	Q8FT90 arabidopsis
13	36	87.8	918	10 Q8RWV9	Q8RWV9 arabidopsis
14	36	87.8	1068	10 Q9AXF7	Q9AXF7 chlamydomon
15	36	87.8	1384	3 Q8XCV2	Q8XCV2 neurospora
16	36	87.8	1559	5 Q86BS0	Q86BS0 drosophila

17	36	87.8	1569	5 Q9VF66	Q9VF66 drosophila
18	36	87.8	1622	5 Q86B80	Q86B80 drosophila
19	35	85.4	146	11 Q8BT98	Q8BT98 mus musculus
20	35	85.4	189	10 Q8LRUS	Q8LRUS tritium ae
21	35	85.4	243	2 Q9FIQ4	Q9FIQ4 thermus the
22	35	85.4	339	13 Q7ZUX2	Q7ZUX2 brachydanio
23	35	85.4	415	5 Q17406	Q17406 caenorhabdi
24	35	85.4	492	15 Q88987	Q88987 viana virus
25	35	85.4	931	12 Q90AP6	Q90AP6 rangiferine
26	35	85.4	1264	10 Q8LSK0	Q8LSK0 zea mays (m
27	35	85.4	1266	5 Q3VER8	Q3VER8 drosophila
28	35	85.4	1426	5 Q8S264	Q8S264 drosophila
29	34	82.9	153	13 Q9W7H7	Q9W7H7 xenopus lae
30	34	82.9	153	13 Q9W7H5	Q9W7H5 squalus aca
31	34	82.9	164	11 Q9D6A0	Q9D6A0 mus musculus
32	34	82.9	164	13 Q9DDV2	Q9DDV2 xenopus lae
33	34	82.9	164	13 Q9DDV3	Q9DDV3 xenopus lae
34	34	82.9	164	13 Q9DDV4	Q9DDV4 xenopus lae
35	34	82.9	165	6 Q9NZAI	Q9NZAI pan troglod
36	34	82.9	165	6 Q9NZAO	Q9NZAO gorilla gor
37	34	82.9	165	6 Q9NZ99	Q9NZ99 pongo pygma
38	34	82.9	182	11 Q8CEB2	Q8CEB2 mus musculus
39	34	82.9	209	4 Q8N811	Q8N811 homo sapien
40	34	82.9	220	13 Q7F339	Q7F339 brachydanio
41	34	82.9	244	13 Q9FMD9	Q9FMD9 xenopus lae
42	34	82.9	244	13 Q8AVH6	Q8AVH6 xenopus lae
43	34	82.9	245	13 Q7T3G8	Q7T3G8 brachydanio
44	34	82.9	252	5 Q7YTD2	Q7YTD2 saccollossu
45	34	82.9	256	13 Q93282	Q93282 brachydanio

## ALIGNMENTS

### RESULT 1

Q7UYG7 PRELIMINARY; PRT; 256 AA.  
AC Q7UYG7  
DT 01-OCT-2003 (TRENBLrel. 25, Created)  
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN RB624.  
OS Rhodopirellula baltica.  
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
OC Planctomycetaceae; Firellula.  
OX NCBI\_TaxID=117;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1;  
RX MEDLINE=22735913; PubMed=12835416;  
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
Schlesner H., Amann R., Reinhardt R.;  
RT "Complete genome sequence of the marine planctomycete Firellula sp.  
strain 1.";  
RT Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
DR EMBL, EX294134; CAD71675.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 256 AA; 27003 MW; 2FAA06B5206E8B2F CRC64;  
Query Match 100.0%; Score 41; DB 16; Length 256;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGP 7

Db 160 RGRPOGP 166

### RESULT 2

Q62106 PRELIMINARY; PRT; 188 AA.  
ID Q62106

Q62106;  
 AC 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Proline-rich salivary protein (Fragment).  
 OS Mus musculus (Mouse).  
 DC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 DC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 DX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=86033799; PubMed=3840480;  
 RA Clements S., Mehansho H., Carlson D.M.;  
 RA "Novel multigene families encoding highly repetitive peptide  
 RT sequences: Sequence analyses of rat and mouse proline-rich protein  
 RT cDNAs.";  
 RL J. Biol. Chem. 260:13471-13477(1985).  
 DR EMBL; M19419; AAA40002.1; -;  
 DR PIR; D29149; D29149. 1  
 FT NON\_TER 1  
 SQ SEQUENCE 188 AA; 18657 MW; A5E98BC5560473FE CRC64;

Query Match 90.2%; Score 37; DB 11; Length 188;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
 :|||||  
 Db 181 QGRPQGP 187

## RESULT 3

Q62103 PRELIMINARY; PRT; 317 AA.  
 AC Q62103;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Proline-rich protein precursor.  
 GN PRP2 OR PRP.  
 OS Mus musculus (Mouse).  
 DC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 DC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 DX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Liver;  
 RX MEDLINE=88273214; PubMed=2839509;  
 RA Ann D.K., Smith M.K., Carlson D.M.;  
 RT "Molecular evolution of the mouse proline-rich protein multigene  
 RT family. Insertion of a long interspersed repeated DNA element.";  
 RL J. Biol. Chem. 263:10887-10893(1988).  
 DR EMBL; M23236; AAA53048.1; -;  
 DR PIR; A28996; A28996.  
 DR MGD; MGI:1932491; Prp2.  
 KW SIGNAL.  
 FT CHAIN 1 15 POTENTIAL.  
 FT SIGNAL 16 317 PROLINE-RICH PROTEIN.  
 SQ SEQUENCE 317 AA; 31719 MW; 019301BE31D73278 CRC64;

Query Match 90.2%; Score 37; DB 11; Length 317;  
 Best Local Similarity 85.7%; Pred. No. 48;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
 :|||||  
 Db 310 QGRPQGP 316

## RESULT 4

Q88189 PRELIMINARY; PRT; 393 AA.  
 AC Q88189;

DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Gag protein (Gag polyprotein) (Fragment).  
 GN GAG.  
 OS Squirrel monkey retrovirus (SMRV-H) (SMRV-HLB).  
 DC Viruses; Retrovirdae; Retroviridae; Betaretrovirus.  
 DX NCBI\_TaxID=11856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B-95-8;  
 RX MEDLINE=95297138; PubMed=7778272;  
 RA Sun R., Grodan E., Shedd D., Bykovsky A.F., Kushnaryov V.M.;  
 RA Grossberg S.E., Miller G.;  
 RA "Transmissible retrovirus in Epstein-Barr virus-producer B95-8  
 RT cells.";  
 RL Virology 209:374-383(1995).  
 DR EMBL; U23805; AAA86516.1; -;  
 DR GO; GO:0019012; C:nucleic acid binding; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; gag\_P24.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR008919; Retrov\_capsid\_N.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00607; Gag\_P24; 1.  
 DR Pfam; PF00098; zf\_CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR SMART; SMO0343; znf\_CCHC; 2.  
 DR PROSITE; PS0158; ZF\_CCHC; 1.  
 KW Core protein; Polyprotein.  
 FT NON\_TER 1  
 FT NON\_TER 393  
 SQ SEQUENCE 393 AA; 43026 MW; F894D5F7CEB244A8 CRC64;

Query Match 90.2%; Score 37; DB 15; Length 393;  
 Best Local Similarity 85.7%; Pred. No. 60;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
 :|||||  
 Db 339 RGRPQGP 345

## RESULT 5

Q8L2K2 PRELIMINARY; PRT; 249 AA.  
 AC Q8L2K2;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DE 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE AcSF.  
 OS Moorella thermoacetica (Clostridium thermoacetum).  
 DC Bacteria; Firmicutes; Clostridia; Thermoanaerobacterales;  
 OC Thermoanaerobacteriaceae; Moorella group; Moorella.  
 DX NCBI\_TaxID=1525;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Loke H.-K., Lindahl P.A.;  
 RT "Identification and Preliminary Characterization of AcSF, a Putative  
 RT Ni-Insertase Used in the Biosynthesis of Acetyl-CoA Synthase from  
 RT Clostridium thermoacetum.";  
 RL J. Inorg. Biochem. 81:0-0(2002).  
 DR EMBL; AF502245; AM22474.1; -;  
 DR InterPro; IPR000707; ATPase\_Para.  
 DR InterPro; IPR001220; Lectin\_legB.  
 DR Pfam; PF00991; Para; 1.  
 DR PROSITE; PSC0307; LECTIN LEGUME BETA; 1.  
 SQ SEQUENCE 249 AA; 27303 MW; 50B208C40625F969 CRC64;

Query Match 87.8%; Score 36; DB 2; Length 249;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

2Y      2 GRPOGP 7
Db      100 GRPOGP 105

RESULT 6
ID Q04154 PRELIMINARY; PRT; 274 AA.
AC Q04154;
JT 01-NOV-1996 (TrEMBLrel. 01, Created)
JT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
JT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
JE Salivary proline-rich protein.
ZN RP15.
DS Rattus norvegicus (Rat).
XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
XC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
XN NCBI_TaxID=10116;
XP [1]
RP SEQUENCE FROM N.A.
IC STRAIN=Sprague-Dawley; TISSUE=Liver;
IX MEDLINE=91257817; PubMed=2045095;
CA Lin H.H., Ann D.K.;
IT "Molecular characterization of rat multigene family encoding proline-
IT rich proteins.";
IL Genomics 10;102-113(1991).
JR EMBL; M64793; AAA42064.1; -.
JR PIR; B39066; B39066.
IQ SEQUENCE 274 AA; 27380 MW; 6A57121F8F07387B CRC64;

Query Match 87.8%; Score 36; DB 11; Length 274;
Best Local Similarity 100.0%; Pred. NO. 64;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y      2 GRPOGP 7
Db      263 GRPOGP 268

RESULT 7
ID Q857B4 PRELIMINARY; PRT; 439 AA.
AC Q857B4;
JT 01-JUN-2002 (TrEMBLrel. 21, Created)
JT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
JT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
JE Hypothetical protein.
ZN OSJNBA0015118.9.
DS Oryza sativa (Rice).
XC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
XC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
XC Ehrhartoideae; Oryzaceae; Oryza.
XN NCBI_TaxID=4530;
XP [1]
RP SEQUENCE FROM N.A.
IC STRAIN=Nipponbare;
RA McCombie W.R.; de la Bastide M.; Spiegel L.; Preston R.; Kirchoff K.;
RA Kuit K.; Nascimento L.; Balija V.; Zutavern T.; Baker J.; Santos L.;
RA Miller B.; Cunniss D.M.; Shah R.; King L.; Katzenberger F.; Muller S.;
RA Bell M.; Yang C.; Dike S.; O'Shaughnessy A.; Palmer L.; Dedhia N.;
RA "Genomic sequence for Oryza sativa, Nipponbare strain, clone
RT OSJNBA0015118, from chromosome 10, complete sequence.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
JR EMBL; AC090482; AAM00980.1; -.
JR Gramene; Q857B4; -.
JR InterPro; IPR008552; DUF834.
JR Pfam; PF05754; DUF834; 1.
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 45162 MW; 3F7F20378E5ABFAC CRC64;

Query Match 87.8%; Score 36; DB 10; Length 439;
Best Local Similarity 85.7%; Pred. NO. 1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2Y      2 GRPOGP 7
Db      67 GRPNCP 73

RESULT 8
ID Q7XFL9 PRELIMINARY; PRT; 439 AA.
AC Q7XFL9;
JT 01-OCT-2003 (TrEMBLrel. 25, Created)
JT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
JT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
ZN OSJNBA0015118.9.
DS Oryza sativa (japonica cultivar-group).
XC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
XC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
XC Ehrhartoideae; Oryzaceae; Oryza.
XN NCBI_TaxID=39947;
XP [1]
RP SEQUENCE FROM N.A.
IC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RA "In-depth view of structure, activity, and evolution of rice
RT chromosome 10.";
RL Science 300;1566-1569(2003).
RN [2]
RP SEQUENCE FROM N.A.
IC STRAIN=cv. Nipponbare;
RA Buell C.R.; Wing R.A.; McCombie W.R.; Messing J.; Yuan Q.;
RA Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017078; AAP53105.1; -.
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 45162 MW; 3F7F20378E5ABFAC CRC64;

Query Match 87.8%; Score 36; DB 10; Length 439;
Best Local Similarity 85.7%; Pred. NO. 1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2Y      1 GRPNCP 7
Db      67 GRPNCP 73

RESULT 9
ID Q8GFF2 PRELIMINARY; PRT; 579 AA.
AC Q8GFF2;
JT 01-MAR-2003 (TrEMBLrel. 23, Created)
JT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
JT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Streptomyces aureofaciens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1894;
XP [1]
RP SEQUENCE FROM N.A.
IC STRAIN=NRRL 2209;
RA Mahishi L.H.; Tripathi G.; Ramachander T.V.N.; Rawal S.K.;
RA "Cloning, molecular analysis and heterologous expression of the
RT poly(3-hydroxybutyrate) synthesizing genes from Streptomyces
RT aureofaciens NRRL 2209.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY032926; AAK53449.1; -.
KW Hypothetical protein.
SQ SEQUENCE 579 AA; 64023 MW; 90BA3409A309789E CRC64;

Query Match 87.8%; Score 36; DB 2; Length 579;
Best Local Similarity 85.7%; Pred. NO. 1.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

2Y 1 RGRPOGP 7
2b 192 RGEPOGP 198

RESULT 10
9RL17 PRELIMINARY; PRT; 603 AA.
AC Q9RL17;
2T 01-MAY-2000 (TrEMBLrel. 13, Created)
2T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
2T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
2E Putative monooxygenase.
2N SC00300 OR SC5G9.09C.
2S Streptomyces coelicolor.
2C Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
2C Streptomycineae; Streptomycetaceae; Streptomyces.
2X NCBI_TaxID=1902;
2N SEQUENCE FROM N.A.
2R STRAIN=A3(2);
2A James K.D., Parkhill J., Barrall B.G., Rajandream M.A.;
2L Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
2R SEQUENCE FROM N.A.
2R STRAIN=A3(2);
2C STRAIN=A3(2);
2X MEDLINE=97000351; PubMed=8843436;
2A Redenbach H., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
2A Kinashi H., Hopwood D.A.;
2R "A set of ordered cosmids and a detailed genetic and physical map for
2T the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
2L Mol. Microbiol. 21:77-96(1996).
2N SEQUENCE FROM N.A.
2R STRAIN=A3(2);
2C STRAIN=A3(2);
2X MEDLINE=21996410; PubMed=12009953;
2A Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
2A Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
2A Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
2A Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
2A Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
2A Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
2A Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
2A Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
2A Hopwood D.A.;
2R "Complete genome sequence of the model actinomycete Streptomyces
2T coelicolor A3(2).";
2L Nature 417:141-147(2002).
2C -1- COFACTOR: FAD (BY SIMILARITY).
2R EMBL; AL939104; CAB55657.1; -.
2R GO; GO:0015036; F:disulfide oxidoreductase activity; IEA.
2R GO; GO:0004497; F:monooxygenase activity; IEA.
2R GO; GO:0006118; P:electron transport; IEA.
2R InterPro; IPR001327; FAD pyr redox.
2R InterPro; IPR001100; Pyr_redox.
2R PRINTS; PR00369; FADPNR.
2R PRINTS; PR00411; FNDKDTASEI.
2R FAD; Flavoprotein; Monooxygenase; Oxidoreductase; Complete proteome.
2X SEQUENCE 603 AA; 66397 MW; 8E95654D5766F544 CRC64;

Query Match 87.8%; Score 36; DB 16; Length 603;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2Y 1 RGRPOGP 7
2b 562 RGRPNCF 568

RESULT 12
Q9FT90 PRELIMINARY; PRT; 917 AA.
ID Q9FT90;
AC Q9FT90;
2T 01-MAR-2001 (TrEMBLrel. 16, Created)
2T 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
2T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

Query Match 87.8%; Score 36; DB 16; Length 803;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 2 GRPOGP 7
2b 195 GRPOGP 200

RESULT 11
Q81ZW2 PRELIMINARY; PRT; 803 AA.
AC Q81ZW2;
2T 01-JUN-2003 (TrEMBLrel. 24, Created)
2T 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
2T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
2E Putative serine/threonine protein kinase.
2N PKN27 OR SAV5424.
2S Streptomyces avermitilis.
2C Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
2C Streptomycineae; Streptomycetaceae; Streptomyces.
2X NCBI_TaxID=33903;
2N SEQUENCE FROM N.A.
2R STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
2R MEDLINE=21477403; PubMed=11572948;
2A Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
2A Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osuno T.,
2A Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
2R "Genome sequence of an industrial microorganism Streptomyces
2T avermitilis: deducing the ability of producing secondary
2R metabolites.";
2L Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
2N SEQUENCE FROM N.A.
2R STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
2R MEDLINE=22608306; PubMed=12692562;
2A Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
2A Sakaki Y., Hattori M., Omura S.;
2R "Complete genome sequence and comparative analysis of the industrial
2T microorganism Streptomyces avermitilis.";
2L Nat. Biotechnol. 21:526-531(2003).
2R EMBL; AP005042; BAC73136.1; -.
2R GO; GO:0005524; P:ATP binding; IEA.
2R GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
2R GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
2R GO; GO:0006118; P:electron transport; IEA.
2R GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
2R InterPro; IPR001128; Cytochrome P450.
2R InterPro; IPR000719; Prot. kinase.
2R InterPro; IPR002290; Ser. Thr. kinase.
2R InterPro; IPR008271; Ser. Thr. kinase.
2R InterPro; IPR001245; Tyr. kinase.
2R Pfam; PF00069; pkinase; 1.
2R ProDom; PD000001; Prot. kinase; 1.
2R SMART; SM00220; S_TKc_1.
2R SMART; SM00219; Ty_Kc_1.
2R PROSITE; PS00086; CYTOCHROME P450; 1.
2R PROSITE; PS00107; PROTEIN KINASE ATP; 1.
2R PROSITE; PS00117; PROTEIN KINASE DOM; 1.
2R PROSITE; PS00108; PROTEIN KINASE ST; 1.
2R Kinase; Serine/threonine-protein kinase; Complete proteome.
2X SEQUENCE 803 AA; 82197 MW; C38C2BE0737FDB66 CRC64;

Query Match 87.8%; Score 36; DB 16; Length 803;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 2 GRPOGP 7
2b 195 GRPOGP 200

RESULT 12
Q9FT90 PRELIMINARY; PRT; 917 AA.
ID Q9FT90;
AC Q9FT90;
2T 01-MAR-2001 (TrEMBLrel. 16, Created)
2T 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
2T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

DE Hypothetical protein.  
GN F8L15.80.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Terry N., Ardiles W., Buyschaert C., Dasseville R.,  
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,  
RA Villarroel R., Gialen J., Van Montagu M., Bancroft I., Mewes H.W.,  
RA Rudd S., Lemcke K., Mayer K.F.X.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL392174; CAC08345.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 917 AA; 104001 MW; 304BC03B584987B1 CRC64;  
Query Match 87.8%; Score 36; DB 10; Length 917;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 RGRPQGP 7  
DB 311 RGNPQGP 317  
RESULT 13  
Q8RW9 PRELIMINARY; PRT; 918 AA.  
AC Q8RW9;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein.  
GN AT5G08450.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,  
RA Davis R.W., Ecker J.R., Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,  
RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Open Reading Frame (ORF) Clones";  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY091073; AM13893.1; -.  
DR EMBL; AY122935; AM67468.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 918 AA; 104152 MW; D224C5B5A0B46A50 CRC64;

Query Match 87.8%; Score 36; DB 10; Length 918;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 RGRPQGP 7  
DB 311 RGNPQGP 317  
RESULT 14  
Q9AXF7 PRELIMINARY; PRT; 1068 AA.  
AC Q9AXF7;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE RNA stability factor MCAI.  
GN MCAI.  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Watson A.T., Gumpel N.J., Ralley L., Purton S.;  
RT "The Chlamydomonas MCAI gene encodes a factor required for petA mRNA stability";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF330231; AAK14341.1; -.  
DR InterPro; IPR002885; PPR.  
DR InterPro; IPR008941; TPR-like.  
DR Pfam; PF01535; PPR; 12.  
DR TIGRFAMs; TIGR00756; PPR; 11.  
SQ SEQUENCE 1068 AA; 109079 MW; 0B289AC32B7E94C3 CRC64;  
Query Match 87.8%; Score 36; DB 10; Length 1068;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GRPQGP 7  
DB 154 GRPQGP 159  
RESULT 15  
Q8XOV2 PRELIMINARY; PRT; 1384 AA.  
AC Q8XOV2;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Conserved hypothetical protein.  
GN 18A7.040.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hohenseel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL670542; CAD21492.1; -.  
DR InterPro; IPR000697; EVH1.  
DR InterPro; IPR004325; Nucleoporin\_FG.  
DR InterPro; IPR00156; Ran\_Bp1.  
DR Pfam; PF03093; Nucleoporin\_FG; 17.  
DR Pfam; PF00638; Ran\_Bp1; 1.  
DR SMART; SM00160; RanBD; 1.

DR PROSITE; PSS0196; RANBD1; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 1384 AA; 142713 MW; 485CB7829339B8B1 CRC64;  
Query Match 87.8%; Score 36; DB 3; Length 1384;  
Best Local Similarity 100.0%; Pred.No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GRPQGP 7  
DB 76 GRPQGP 81  
Search completed: April 6, 2004, 16:14:38  
Job time : 26.3947 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 39.1869 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-3

Perfect score: 41

Sequence: 1 RGRPQP 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	7	4 AAB48773	Aab48773 Human sal
2	41	100.0	8	4 AAB48774	Aab48774 Human sal
3	41	100.0	9	4 AAB48775	Aab48775 Human sal
4	41	100.0	10	4 AAB48776	Aab48776 Human sal
5	41	100.0	17	4 AAB48783	Aab48783 Human sal
6	41	100.0	132	4 AAB38848	Aab38848 Peptide #
7	41	100.0	132	4 AAM32323	Aam32323 Peptide #
8	41	100.0	132	4 AAM72058	Aam72058 Human bon
9	41	100.0	132	4 AAM59494	Aam59494 Human bra
10	41	100.0	132	4 ABG53742	Abg53742 Human liv
11	41	100.0	132	5 ABG41873	Abg41873 Human pep
12	41	100.0	149	6 ABR57423	Abr57423 Human NOV
13	41	100.0	154	6 ABR56769	Abr56769 Human sec
14	41	100.0	166	6 ADAB3798	Adab3798 Human PRH
15	41	100.0	166	7 ADC98216	Adc98216 Human sal
16	37	90.2	19938	6 ABP76679	Abp76679 Streptomy
17	36	87.8	44	2 AAWC3557	Aawc3557 Human pro
18	36	87.8	45	1 AAO93320	Aao93320 P-C Gene.
19	36	87.8	51	4 AAO08948	Aao08948 Human pol
20	36	87.8	58	3 AAY91436	Aay91436 Human sec
21	36	87.8	62	4 AAV65614	Aav65614 Propionib
22	36	87.8	62	6 ABM64852	Abm64852 Propionib
23	36	87.8	62	6 ABM63133	Abm63133 Propionib
24	36	87.8	63	4 AAO08425	Aao08425 Human pol
25	36	87.8	64	2 AAY29533	Aay29533 Human lun

ALIGNMENTS

RESULT 1  
AAB48773  
ID AAB48773 standard; peptide; 7 AA.  
XX  
AC AAB48773;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human saliva PRP-1 fragment (residues 106-112), SEQ ID NO:3.  
XX  
KW Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX  
OS Homo sapiens.  
XX  
PN WO200069890-A1.  
XX  
PD 23-NOV-2000.  
XX  
PF 11-MAY-2000; 2000WO-SE000930.  
XX  
PR 17-MAY-1999; 99SE-00001773.  
XX  
PA (STRO/) STROEMBERG N.  
XX (JOHA/) JOHANSSON I.  
PI Stroemberg N, Johansson I;  
XX WPI; 2001-03:923/04.  
DR  
PT New oligopeptides comprising 2 arginine residues from degradation of  
PT proline-rich proteins, useful for preventing dental caries.  
PS  
CC Claim 4; Page 24; 36pp; English.  
CC  
CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
CC AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes. With certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

Aab44432 Human lun  
Aae13773 Human lun  
Add66382 Human lun  
Ade87636 Human lun  
Aab4508 Human sec  
Aab64507 Gene 25 h  
Abg19115 Novel hum  
Abb79303 Human ova  
Aam96563 Human rep  
Abb96567 Human tes  
Aao7824 Human pol  
Abj03746 Human ova  
Abp75970 Human GEN  
Abp76136 Human GEN  
Abj03765 Human ova  
Aao30407 Human sec  
Abbi1118 Human sec  
Aao30408 Human sec  
Add47240 Rat Prote  
Add48703 Rat Prote

CC derived oligopeptides of the invention

XX Sequence 7 AA;

Query Match 100.0%; Score 41; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQGP 7  
Db 1 RGRPQGP 7

RESULT 2

AAB48774 ID AAB48774 standard; peptide; 8 AA.

XX AAB48774;

AC AC

XX 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 106-113), SEQ ID NO:4.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX Homo sapiens.

OS WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO//) STROEMBERG N.

PA (JOHA//) JOHANSSON I.

XX Stroemberg N, Johansson I;

PI WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of

PT proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
CC derived oligopeptides of the invention

XX Sequence 8 AA;

Query Match 100.0%; Score 41; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQGP 7  
Db 1 RGRPQGP 7

RESULT 3

AAB48775 ID AAB48775 standard; peptide; 9 AA.

XX AAB48775;

AC AC

XX 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 106-114), SEQ ID NO:5.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX Homo sapiens.

OS WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO//) STROEMBERG N.

PA (JOHA//) JOHANSSON I.

XX Stroemberg N, Johansson I;

PI WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of

PT proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
CC derived oligopeptides of the invention

XX Sequence 9 AA;

Query Match 100.0%; Score 41; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQGP 7  
Db 1 RGRPQGP 7

RESULT 4

AAB48776 ID AAB48776 standard; peptide; 10 AA.

XX AAB48776;

AC AC

XX 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 106-115), SEQ ID NO:6.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX Homo sapiens.  
 OS WO200069890-A1.  
 PN XX  
 XX 23-NOV-2000.  
 PD XX  
 XX 11-MAY-2000; 2000WO-SE000930.  
 PF XX  
 XX 17-MAY-1999; 99SE-00001773.  
 PR XX  
 XX (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX Stroemberg N, Johansson I;  
 PI WPI; 2001-031923/04.  
 XX  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX  
 XX Claim 4; Page 24; 36pp; English.  
 PS The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX  
 XX Sequence 10 AA;  
 SQ  
 Query Match 100.0%; Score 41; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. NO. 1.2; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0;  
 QY 1 RGRPQGP 7  
 Db |||||  
 1 RGRPQGP 7  
 RESULT 5  
 AAB48783  
 ID AAB48783 standard; peptide; 17 AA.  
 XX  
 AC AAB48783;  
 XX  
 XX 09-MAR-2001 (first entry)  
 DT Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.  
 XX  
 DE Human; PRP-1; proline-rich protein; saliva; dental caries;  
 KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200069890-A1.  
 PN XX  
 XX 23-NOV-2000.  
 PD XX  
 XX 11-MAY-2000; 2000WO-SE000930.  
 PF XX  
 XX 17-MAY-1999; 99SE-00001773.  
 PR XX

XX (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX Stroemberg N, Johansson I;  
 PI WPI; 2001-031923/04.  
 XX  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX  
 XX Claim 2; Page 24; 36pp; English.  
 PS The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX  
 XX Sequence 17 AA;  
 SQ  
 Query Match 100.0%; Score 41; DB 4; Length 17;  
 Best Local Similarity 100.0%; Pred. NO. 2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPQGP 7  
 Db |||||  
 8 RGRPQGP 14  
 RESULT 6  
 ABB38848  
 ID ABB38848 standard; peptide; 132 AA.  
 XX  
 AC ABB38848;  
 XX  
 XX 04-FEB-2002 (first entry)  
 DT Peptide #6354 encoded by human foetal liver single exon probe.  
 DE Human; foetal liver; gene expression; single exon nucleic acid probe.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200157277-A2.  
 PN  
 XX 09-AUG-2001.  
 PD  
 XX 30-JAN-2001; 2001WO-US000669.  
 XX  
 XX 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-483447/52.  
 XX  
 XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT

PT gene expression in human fetal liver.

PS Claim 27; SEQ ID NO 31483; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring  
CC human gene expression in a sample derived from human foetal liver. The  
CC single exon nucleic acid probes may be used for predicting, measuring and  
CC displaying gene expression in samples derived from human fetal liver. The  
CC present sequence is a peptide encoded by a single exon nucleic acid probe  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 132 AA;

Query Match 100.0%; Score 41; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
Db 88 RGRPQGP 94

RESULT 7  
AAM32323  
ID AAM32323 standard; protein; 132 AA.

XX AAM32323;

XX 17-OCT-2001 (first entry)

DE Peptide #6360 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;  
XX Genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488997/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human placenta.

XX Claim 27; SEQ ID NO 32592; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs;  
XX see AAI31315-AA157546). The present sequence is a peptide encoded by one  
XX such probe. The probes are useful for producing a microarray for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from human placenta. The probes are useful for antenatal diagnosis of  
XX human genetic disorders

XX Sequence 132 AA;

Query Match 100.0%; Score 41; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
Db 88 RGRPQGP 94

RESULT 8  
AAM72058  
ID AAM72058 standard; protein; 132 AA.

XX AAM72058;

XX 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32364.

XX Human; bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000668.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488990/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human bone marrow.

XX Example 4; SEQ ID NO 32364; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is a  
XX protein encoded by one of the probes of the invention

XX Sequence 132 AA;

Query Match 100.0%; Score 41; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
Db 88 RGRPQGP 94

RESULT 9  
AAM59494  
ID AAM59494 standard; protein; 132 AA.

XX AAM59494;

XX

```

DT 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 31599.
DE Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
KW Homo sapiens.
XX WO200157275-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000667.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX Example 4; SEQ ID NO 31599; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX SQ Sequence 132 AA;
XX Query Match 100.0%; Score 41; DB 4; Length 132;
XX Best Local Similarity 100.0%; Pred. NO. 14;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 RGRPQGP 7
XX DB 88 RGRPQGP 94
XX RESULT 10
XX ABG53742
XX ID ABG53742 standard; peptide; 132 AA.
XX AC ABG53742;
XX 25-FEB-2003 (first entry)
XX Human liver peptide, SEQ ID No 32390.
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX hypercholesterolaemia; coronary heart disease.
XX Homo sapiens.
XX WO200157273-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000664.

XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488898/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX Claim 27; SEQ ID NO 32390; 658pp; English.
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG5930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 132 AA;
XX Query Match 100.0%; Score 41; DB 4; Length 132;
XX Best Local Similarity 100.0%; Pred. NO. 14;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 RGRPQGP 7
XX DB 88 RGRPQGP 94
XX RESULT 11
XX ABG41873
XX ID ABG41873 standard; peptide; 132 AA.
XX AC ABG41873;
XX 19-AUG-2002 (first entry)
XX Human peptide encoded by genome-derived single exon probe SEQ ID 31538.
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX Homo sapiens.
XX WO200186003-A2.
XX 15-NOV-2001.
XX 30-JAN-2001; 2001WO-US000665.

```

XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WFI; 2002-114193/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
XX  
XX Claim 27; SEQ ID NO 31538; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WFO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 132 AA;  
SQ  
Query Match 100.0%; Score 41; DB 5; Length 132;  
Best Local Similarity 100.0%; Freq. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRPQGP 7  
DB 88 RGRPQGP 94  
|||||  
RESULT 12  
ABR57423  
ID ABR57423 standard; protein; 149 AA.

XX ABR57423;  
XX 15-SEP-2003 (first entry)  
XX Human NOV7 protein SEQ ID NO:24.  
XX  
XX Human; NOVX; cytostatic; cardiant; antiinflammatory; immunosuppressive;  
KW antiallergic; haemostatic; anti-HIV; antidiabetic; antiarteriosclerotic;  
KW anorectic; antiasthmatic; nephrotropic; antiarthritic; hepatoclerotic;  
KW neuroprotective; nontropic; antibacterial; virucide; antiparasitic;  
KW relaxant; anticonvulsant; hypotensive; vasotropic; vaccine; cancer;  
KW vulnary; angiogenic; antiangiogenic; gene therapy; vaccine; inflammation;  
KW cardiomyopathy; atherosclerosis; hypertension; diabetes; asthma;  
KW autoimmune disorder; allergy; blood disorder; AIDS; obesity; asthma;  
KW acquired immunodeficiency syndrome; nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; Parkinson's disease; goitre; infection; stroke;  
KW muscular dystrophy; epilepsy; wasting disorder; chromosome 12.  
XX  
XX Homo sapiens.  
XX  
XX WO200294870-A2.  
XX  
XX 28-NOV-2002.  
XX  
XX 02-NOV-2001; 2001WO-US051580.  
XX  
XX 02-NOV-2000; 2000US-0245291P.  
PR 02-NOV-2000; 2000US-0245317P.  
PR 07-NOV-2000; 2000US-0245662P.  
PR 08-NOV-2000; 2000US-0246871P.  
PR 26-JAN-2001; 2001US-0264389P.  
PR 26-JAN-2001; 2001US-0264423P.  
PR 29-JAN-2001; 2001US-0264799P.  
XX (CURA-) CURAGEN CORP.  
XX  
XX Grosse WM, Macdougall JR, Smithson G, Millet I, Stone DJ;  
PI Gunther E, Ellerman K, Alsobrook JP, Lepley DM, Burgess CE;  
PI Sytek KA, Edinger SR, Gangoli EA, Gorman L, Taupier RJ, Li L;  
PI Guo X, Fernandes ER, Vernet CAM, Tchernev VT, Caeman SJ, Shenoy S;  
PI Mishra V, Furtak K, Baumgartner JC, Colman SD;  
XX WFI; 2003-140359/13.  
DR N-PSDB; ACF03558.  
XX  
XX New NOVX polypeptide useful for preventing or treating NOVX-associated  
PT disorders, e.g. cancer, cardiomyopathy, atherosclerosis or diabetes, and  
PT in chromosome mapping, tissue typing or pharmacogenomics.  
XX  
XX Claim 1; Page 59; 345pp; English.  
XX  
XX ACP03547 to ACP03570 encode the human NOVX proteins (I) given in ABR57412  
CC to ABR57435. (I) Have cytostatic, cardiant, antiinflammatory, nontropic,  
CC immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic,  
CC antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, virucide,  
CC antiarthritic, hepatoclerotic, neuroprotective, antibacterial, relaxant,  
CC antiparasitic, anticonvulsant, hypotensive, vasotropic, antiparkinsonian,  
CC vulnary, angiogenic and antiangiogenic activities, and can be used in  
CC gene therapy and vaccines. The NOVX polypeptides and their antibodies can  
CC be used to determine the presence or absence of (I) in a sample. The NOVX  
CC polypeptides, polynucleotides encoding them, and antibodies against them,  
CC are useful in manufacturing a medicament for treating or preventing a  
CC syndrome associated with a NOVX-associated disorder such as hypertension,  
CC cardiomyopathy, atherosclerosis, cancer, diabetes, asthma, inflammation,  
CC autoimmune disorders, allergies, blood disorders, obesity, acquired  
CC immunodeficiency syndrome (AIDS), immunoglobulin (Ig) A nephropathy,  
CC cirrhosis, arthritis, Alzheimer's disease, Parkinson's disease, goitre,  
CC infections (e.g. bacterial, viral, parasitic), stroke, muscular  
CC dystrophy, epilepsy, and other wasting disorders associated with chronic  
CC diseases. ACP03571 to ACP03644 represent PCR primers and probes for NOVX  
CC sequence, which are used in an example from the present invention  
XX

SQ Sequence 149 AA;  
 Query Match 100.0%; Score 41; DB 6; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPQGP 7  
 |||||  
 Db 105 RGRPQGP 111  
 |||||  
 RESULT 13  
 ABR56769  
 ID ABR56769 standard; protein; 154 AA.  
 XX  
 AC ABR56769;  
 XX  
 DT 30-JUL-2003 (first entry)  
 XX  
 DE Human secreted protein SECP-44 SEQ ID NO:44.  
 XX  
 KW Human; secreted protein; SECP; cytostatic; antiarteriosclerotic;  
 KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;  
 KW antiallergic; antiinflammatory; thyromimetic; gene therapy; cancer;  
 KW cell proliferative disorder; atherosclerosis; neurological disorder;  
 KW epilepsy; Huntington's disease; stroke; immune disorder; allergy; AIDS;  
 KW inflammatory disorder; developmental disorder; hypothyroidism;  
 KW Cushing's syndrome; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W02003016506-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 15-AUG-2002; 2002WO-US027143.  
 XX  
 PR 17-AUG-2001; 2001US-0313249P.  
 PR 24-AUG-2001; 2001US-0314752P.  
 PR 07-SEP-2001; 2001US-0317818P.  
 PR 21-SEP-2001; 2001US-0317824P.  
 PR 24-SEP-2001; 2001US-0324040P.  
 PR 02-NOV-2001; 2001US-0324586P.  
 PR 28-NOV-2001; 2001US-0343980P.  
 PR 13-FEB-2002; 2001US-0334229P.  
 PR 06-MAR-2002; 2002US-0357002P.  
 PR 19-MAR-2002; 2002US-0362439P.  
 PR 30-APR-2002; 2002US-0376988P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Tang TY, Warren BA, Gietzen KJ, Lal PG, Yue H, Honchell CD;  
 PI Lehr-Watson PM, Buford N, Xu Y, Baughn MR, Dugan BM, Tran UK;  
 PI Lee EA, Forsythe IJ, Richardson TM, Lee S, Thangavelu K, Yue H;  
 PI Emerling ML, Walia NK, Azimzai Y, Sanjanwala B, Hafalia AJA;  
 PI Borowsky ML, Nguyen DB, Ison CH, Astronoff A, Ding L, Lee SY;  
 PI Becha SD, Ramkumar J, Gandhi AR, Jin P, Fu GK, Swarnakar A;  
 XX  
 DR WPI; 2003-278569/27.  
 DR N-PSDB; ACC79069.  
 XX  
 PT New human secreted proteins (SECP), useful for diagnosing, treating and  
 PT preventing diseases or conditions associated with the aberrant SECP  
 PT expression e.g. cancer, AIDS, atherosclerosis, Huntington's disease,  
 PT stroke, infections.  
 XX  
 PS Claim 1; Page 222; 286pp; English.  
 XX  
 CC ACC79026 to ACC79105 encode the human secreted proteins (I) given in  
 CC ABR56726 to ABR56805, designated SECP-1 to SECP-80. SECP sequences can  
 CC have cytostatic, antiarteriosclerotic, anticonvulsant, antiinflammatory,  
 CC nootropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic and

CC thymomimetic activities, and can be used in gene therapy. The SECP  
 CC proteins and polynucleotides can be used in diagnosing, treating and  
 CC preventing diseases or conditions associated with the decreased  
 CC expression or overexpression of SECP, such as cell proliferative (e.g.  
 CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's  
 CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and  
 CC developmental (e.g. hypothyroidism, Cushing's syndrome) disorders, or  
 CC infections. They are also useful in assessing the effects of exogenous  
 CC compounds on the expression of nucleic acid and amino acid sequences of  
 CC SECP. The SECP or its fragments are useful in screening compounds for  
 CC effectiveness as agonist or antagonist of the polypeptides, or in  
 CC altering the expression of the target polynucleotide and compounds that  
 CC specifically bind to or modulate the activity of the polypeptide  
 XX  
 SQ Sequence 154 AA;  
 Query Match 100.0%; Score 41; DB 6; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPQGP 7  
 |||||  
 Db 110 RGRPQGP 116  
 |||||  
 RESULT 14  
 ADA83798  
 ID ADA83798 standard; protein; 166 AA.  
 XX  
 AC ADA83798;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Human PRH2 protein.  
 XX  
 KW human; marker; expressed sequence tag; EST; arabidopsis; tumour;  
 KW stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;  
 KW vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W02002103028-A2.  
 XX  
 PD 27-DEC-2002.  
 XX  
 PF 30-MAY-2002; 2002WO-IB004189.  
 XX  
 PR 30-MAY-2001; 2001US-0293999P.  
 PR 22-OCT-2001; 2001US-0330457P.  
 PR 19-FEB-2002; 2002US-0357144P.  
 XX  
 PA (BIOM-) BIOMEDICAL CENT.  
 XX  
 PI Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;  
 PI WPI; 2003-175241/17.  
 DR N-PSDB; ADA83797.  
 XX  
 PT Determining if a nucleic acid is a marker for a phenotype/cell type of  
 PT interest, by global comparison of expressed sequence tags known to be  
 PT expressed in the phenotype/cell type with all ESTs expressed in normal  
 PT tissue.  
 XX  
 PS Claim 29; Page 191-192; 516pp; English.  
 XX  
 CC The invention relates to a novel method for determining if a nucleic acid  
 CC is a marker for a predetermined phenotype/cell type of interest from a  
 CC biological species. The method comprises performing a global comparison  
 CC of a group of expressed sequence tags (ESTs) known to be expressed in the  
 CC phenotype/cell type of interest with all ESTs expressed in normal tissue  
 CC in order to identify ESTs that are preferentially expressed in the  
 CC phenotype/cell of interest. A method of the invention is useful for  
 CC determining whether a nucleic acid is a marker for a predetermined

CC phenotype or cell type of interest from a biological species, preferably  
CC Arabidopsis or human. The cell type of interest is an abnormal cell such  
CC as a tumour cell, and the predetermined phenotype is a stress-induced  
CC phenotype such as hyperosmotic stress or high salt conditions. A method  
CC of the invention is also useful for determining the progression of colon  
CC cancer in a human, for detecting a tumour cell, and for regulating or  
CC preventing the growth of a tumour cell. An antibody of the invention is  
CC useful for detecting the absence or presence of peptides encoded by  
CC tumour-associated markers. A polypeptide of the invention is useful as an  
CC immunogen for vaccinating an animal. The present sequence represents a  
CC tumour-associated antigen of the invention.  
XX  
SQ Sequence 166 AA;

Query Match 100.0%; Score 41; DB 6; Length 166;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRPQGP 7  
Db 122 RGRPQGP 128  
|||||

RESULT 15  
ADC98216  
ID ADC98216 standard; protein; 166 AA.  
XX AC ADC98216;  
XX DT 01-JAN-2004 (first entry)  
XX DE Human salivary acidic proline-rich phosphoprotein (PRP).  
XX Human; salivary acidic proline-rich phosphoprotein; PRP; autoantigen;  
XX atopic dermatitis-inducing protein; salivary gland; IgE autoantibody;  
XX immunoglobulin E; mast cell activation; basophil activation; diagnosis;  
XX risk assessment; sensitisation remedy; dermatological; anti-allergic;  
XX anti-inflammatory.  
XX Homo sapiens.  
XX OS  
XX WO2003084991-A1.  
XX 16-OCT-2003.  
XX 04-APR-2003; 2003WO-JP004325.  
XX 08-APR-2002; 2002JP-00105425.  
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX Hide M, Yamamoto S, Tanaka T, Koro O;  
XX WPI; 2003-833567/77.  
XX N-PSDB; ADC98215.  
XX Atopic dermatitis-inducing proteins, applicable in diagnosis of including  
XX risk of onset, and in developing sensitization remedies for the disease.  
XX Claim 4; SEQ ID NO 2; 43pp; Japanese.  
XX The invention relates to the human atopic dermatitis-inducing proteins,  
XX salivary acidic proline-rich phosphoprotein (PRP; ADC98216) and prolactin  
XX -inducible protein (PIP; ADC98218), and their post-translationally  
XX modified forms. These proteins are secreted by salivary or sweat glands  
XX and bind to IgE autoantibodies, thereby activating mast cells and  
XX basophils. The invention also relates to antigenic peptide fragments of  
XX PRP or PIP; antibodies which bind to PRP or PIP; methods for diagnosing  
XX atopic dermatitis or for determining whether an individual is at risk of  
XX developing atopic dermatitis by determining the presence of PRP- or PIP-  
XX specific antibodies or immune complexes, or by quantifying histamine  
XX release; and sensitisation remedies for atopic dermatitis containing PRP  
XX and/or PIP or their peptides as the active ingredient. PRP, PIP and their

CC antibodies are useful in diagnosing atopic dermatitis, or for determining  
CC whether an individual is at risk of developing atopic dermatitis. They  
CC are also useful in developing sensitisation remedies for the treatment of  
CC atopic dermatitis. The present sequence represents the specifically  
CC claimed human salivary acidic proline-rich phosphoprotein (PRP).  
XX  
SQ Sequence 166 AA;

Query Match 100.0%; Score 41; DB 7; Length 166;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRPQGP 7  
Db 122 RGRPQGP 128  
|||||

Search completed: April 6, 2004, 16:06:37  
Job time : 40.1869 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 27.4766 Seconds  
(without alignments)  
66,909 Million cell updates/sec

Title: US-10-009-709-3

Perfect score: 41

Sequence: 1 RGRPQGP 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 107172 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	132	9	US-09-864-761-43644
2	41	100.0	166	14	US-10-157-031-80
3	37	90.2	440	12	US-10-425-114-39512
4	37	90.2	19723	15	US-10-084-846A-5
5	36	87.8	58	9	US-09-739-254-157
6	36	87.8	58	9	US-09-904-615-157
7	36	87.8	58	14	US-10-054-388-157
8	36	87.8	58	14	US-10-055-098-157
9	36	87.8	64	9	US-09-738-973-74
10	36	87.8	64	9	US-09-854-133-74
11	36	87.8	64	14	US-10-144-649A-74
12	36	87.8	73	15	US-10-108-260A-3869
13	36	87.8	75	12	US-10-001-885-100
14	36	87.8	77	10	US-09-764-891-5221
15	36	87.8	89	13	US-10-001-835-188

16	36	87.8	130	15	US-10-108-260A-2808	Sequence 2808, Ap
17	36	87.8	135	13	US-10-001-835-207	Sequence 207, App
18	36	87.8	151	15	US-10-108-260A-4055	Sequence 4055, Ap
19	36	87.8	175	12	US-10-276-774-1488	Sequence 1488, Ap
20	36	87.8	274	9	US-09-850-887-4	Sequence 4, Appli
21	36	87.8	803	14	US-10-156-761-12958	Sequence 12958, A
22	36	87.8	958	13	US-10-095-929-8	Sequence 8, Appli
23	36	87.8	960	13	US-10-095-929-3	Sequence 3, Appli
24	35	85.4	30	10	US-09-892-877-307	Sequence 307, App
25	35	85.4	30	10	US-09-948-793-320	Sequence 320, App
26	35	85.4	183	12	US-10-425-114-70816	Sequence 70816, A
27	35	85.4	548	14	US-10-307-019-9	Sequence 9, Appli
28	35	85.4	761	12	US-10-425-114-49136	Sequence 49136, A
29	35	85.4	819	16	US-10-389-566-480	Sequence 480, App
30	35	85.4	19608	15	US-10-084-846A-8	Sequence 8, Appli
31	34	82.9	37	9	US-09-864-761-34464	Sequence 34464, A
32	34	82.9	89	12	US-10-424-599-185980	Sequence 185980, A
33	34	82.9	149	12	US-10-424-599-171613	Sequence 171613, A
34	34	82.9	163	12	US-10-424-599-189117	Sequence 189117, A
35	34	82.9	209	15	US-10-108-260A-4370	Sequence 4370, Ap
36	34	82.9	308	12	US-10-425-114-61262	Sequence 61262, A
37	34	82.9	331	14	US-10-156-761-13112	Sequence 13112, A
38	34	82.9	332	12	US-10-425-114-59820	Sequence 59820, A
39	34	82.9	582	15	US-10-108-260A-4485	Sequence 4485, Ap
40	34	82.9	1159	9	US-09-735-995-2	Sequence 2, Appli
41	34	82.9	1159	9	US-09-735-995-4	Sequence 4, Appli
42	34	82.9	1159	9	US-09-119-855-10	Sequence 10, Appli
43	34	82.9	1159	14	US-10-000-151B-3	Sequence 3, Appli
44	34	82.9	1159	14	US-10-193-692-5	Sequence 5, Appli
45	34	82.9	1159	14	US-10-185-867-10	Sequence 10, Appli

#### ALIGNMENTS

#### RESULT 1

```
US-09-864-761-43644
; Sequence 43644, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, David R.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 43644  
LENGTH: 132  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC006518.17  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77  
OTHER INFORMATION: EST HUMAN HIT: BF088785.1, EVALUE 1.00e-01  
OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02  
US-09-864-761-43644

Query Match 100.0%; Score 41; DB 9; Length 132;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
Db 88 RGRPQGP 94

RESULT 2  
US-10-157-031-80  
Sequence 80, Application US/10157031  
Publication No. US20030108890A1  
GENERAL INFORMATION:  
APPLICANT: Baranova, A. V.  
APPLICANT: Yankovsky, N. K.  
APPLICANT: Kozlov, A. P.  
APPLICANT: Lobashev, A. V.  
APPLICANT: Kravkovskaya, I. L.  
TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences  
FILE REFERENCE: 2760-103  
CURRENT APPLICATION NUMBER: US/10/157,031  
CURRENT FILING DATE: 2002-05-30  
NUMBER OF SEQ ID NOS: 415  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 80  
LENGTH: 166  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-157-031-80

Query Match 100.0%; Score 41; DB 14; Length 166;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
Db 122 RGRPQGP 128

RESULT 3  
US-10-425-114-39512  
Sequence 39512, Application US/10425114

Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 39512  
LENGTH: 440  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700240895\_FLI.pep  
US-10-425-114-39512

Query Match 90.2%; Score 37; DB 12; Length 440;  
Best Local Similarity 85.7%; Pred. No. 3.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
Db 420 RGRPQGP 426

RESULT 4  
US-10-084-846A-5  
Sequence 5, Application US/10084846A  
Publication No. US20040006026A1  
GENERAL INFORMATION:  
APPLICANT: WEITNAUER, GABRIELE  
APPLICANT: MUEHLEWEG, AGNES  
APPLICANT: TREFFER, AXEL  
APPLICANT: BECHTHOLD, ANDREAS  
TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
FILE REFERENCE: 1974-005  
CURRENT APPLICATION NUMBER: US/10/084,846A  
CURRENT FILING DATE: 2003-02-25  
PRIOR APPLICATION NUMBER: PCT/EP01/09815  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: DE 101 09 166.4  
PRIOR FILING DATE: 2001-02-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 3.2  
SEQ ID NO 5  
LENGTH: 19723  
TYPE: PRT  
ORGANISM: Streptomyces viridochromogenes  
FEATURE:  
OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.  
OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.  
US-10-084-846A-5

Query Match 90.2%; Score 37; DB 15; Length 19723;  
Best Local Similarity 85.7%; Pred. No. 1.1e+04;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7  
Db 10188 RGRPQGP 10194

RESULT 5  
US-09-739-254-157  
Sequence 157, Application US/09739254  
Patent No. US20010021700A1  
GENERAL INFORMATION:

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/739,254
; CURRENT FILING DATE: 2000-12-19
; EARLIER APPLICATION NUMBER: 09/511,554
; EARLIER FILING DATE: 2000-02-23
; EARLIER APPLICATION NUMBER: PCT/US99/19330
; EARLIER FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 60/097,917
; EARLIER FILING DATE: 1998-08-25
; EARLIER APPLICATION NUMBER: 60/098,634
; EARLIER FILING DATE: 1998-08-31
; SOFTWARE: PatentIn Ver. 2.0
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 157
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-254-157

```

```

Query Match      87.8%; Score 36; DB 9; Length 58;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 GRPQGP 7
DB      36 GRPQGP 41

```

## RESULT 6

```

US-09-904-615-157
; Sequence 157, Application US/09904615
; Publication No. US2002026040A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/904,615
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 157
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-904-615-157

```

```

Query Match      87.8%; Score 36; DB 9; Length 58;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 GRPQGP 7
DB      36 GRPQGP 41

```

## RESULT 7

```

US-10-054-988-157
; Sequence 157, Application US/10054988
; Publication No. US20030087341A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/10/054,988
; CURRENT FILING DATE: 2002-01-25

```

```

; PRIOR APPLICATION NUMBER: 09/904,615
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 157
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-988-157

```

```

Query Match      87.8%; Score 36; DB 14; Length 58;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 GRPQGP 7
DB      36 GRPQGP 41

```

## RESULT 8

```

US-10-055-098-157
; Sequence 157, Application US/10055098
; Publication No. US20030139954A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/10/055,098
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: EARLIER FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US99/19330
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 157
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-055-098-157

```

```

Query Match      87.8%; Score 36; DB 14; Length 58;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 GRPQGP 7
DB      36 GRPQGP 41

```

## RESULT 9

```

US-05-738-973-74
; Sequence 74, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph

```

```

; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-74

```

Query Match 87.8%; Score 36; DB 9; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7  
 |||||  
 Db 23 GRPQGP 28

```

RESULT 10
US-09-854-133-74
; Sequence 74, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-74

```

Query Match 87.8%; Score 36; DB 9; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7  
 |||||  
 Db 23 GRPQGP 28

```

RESULT 11
US-10-144-649A-74
; Sequence 74, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21

```

```

; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-144-649A-74

```

Query Match 87.8%; Score 36; DB 14; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7  
 |||||  
 Db 23 GRPQGP 28

```

RESULT 12
US-10-108-260A-3869
; Sequence 3869, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560a1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3869
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3869

```

Query Match 87.8%; Score 36; DB 15; Length 73;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7  
 |||||  
 Db 32 GRPQGP 37

```

RESULT 13
US-10-001-885-100
; Sequence 100, Application US/10001885
; Publication No. US20040058319A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caifferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pri
; FILE REFERENCE: DEX-0279
; CURRENT APPLICATION NUMBER: US/10/001,885
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,061
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/253,257
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 100
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-885-100

```

Query Match 87.8%; Score 36; DB 12; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 97;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQP 7  
Db 32 GRPQP 37

## RESULT 14

US-09-764-891-5221  
; Sequence 5221, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5221  
; LENGTH: 77  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-891-5221

Query Match 97.8%; Score 36; DB 10; Length 77;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQP 7  
Db 36 GRPQP 41

## RESULT 15

US-10-001-835-188  
; Sequence 188, Application US/10001835  
; Publication No. US20020160387A1  
; GENERAL INFORMATION:  
; APPLICANT: Salceda, Susana  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Cafferkey, Robert  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro  
; FILE REFERENCE: DEX-0277  
; CURRENT APPLICATION NUMBER: US/10/001,835  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/249,997  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 228  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 188  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-001-835-188

Query Match 87.8%; Score 36; DB 13; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQP 7  
Db 32 GRPQP 37

Search completed: April 6, 2004, 17:05:52  
Job time : 28.4766 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 10.271 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-3

Perfect score: 41

Sequence: 1 RGRPQGP 7

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA:\*

2: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pep.\*

7: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	92.7	423	4	US-09-489-039A-7834
2	37	90.2	446	4	US-09-252-991A-27110
3	36	87.8	58	4	US-09-904-615-157
4	36	87.8	64	4	US-09-370-838-74
5	36	87.8	73	1	US-08-355-888A-11
6	36	87.8	73	2	US-08-693-697-11
7	36	87.8	73	3	US-08-693-696-11
8	36	87.8	73	4	US-09-357-914-11
9	36	87.8	132	4	US-09-252-991A-22177
10	36	87.8	139	4	US-09-252-991A-27546
11	36	87.8	239	4	US-09-252-991A-26949
12	36	87.8	243	4	US-09-252-991A-16906
13	36	87.8	473	4	US-09-252-991A-21690
14	36	87.8	958	2	US-08-640-389A-8
15	36	87.8	958	4	US-08-618-957A-8
16	36	87.8	960	1	US-08-355-888A-8
17	36	87.8	960	2	US-08-588-190-3
18	36	87.8	960	2	US-08-693-697-8
19	36	87.8	960	2	US-08-640-389A-3
20	36	87.8	960	3	US-08-693-696-8
21	36	87.8	960	4	US-08-618-957A-3
22	36	87.8	960	4	US-09-357-914-8
23	35	85.4	149	4	US-09-252-991A-32644
24	35	85.4	384	4	US-09-252-991A-31787
25	35	85.4	391	4	US-09-252-991A-20316
26	35	85.4	456	4	US-09-252-991A-26668
27	35	85.4	618	4	US-09-252-991A-29306

28 35 85.4 651 4 US-09-252-991A-19760 Sequence 19760, A  
29 34 82.9 270 4 US-09-252-991A-32606 Sequence 32606, A  
30 34 82.9 386 4 US-09-252-991A-17018 Sequence 17018, A  
31 34 82.9 392 4 US-09-252-991A-27798 Sequence 27798, A  
32 34 82.9 399 4 US-09-252-991A-25098 Sequence 25098, A  
33 34 82.9 474 4 US-09-252-991A-28084 Sequence 28084, A  
34 34 82.9 1159 2 US-08-956-242-13 Sequence 13, Appl  
35 34 82.9 1159 3 US-09-351-215-13 Sequence 13, Appl  
36 34 82.9 1159 3 US-09-226-012-2 Sequence 2, Appl  
37 34 82.9 1159 3 US-09-226-012-4 Sequence 4, Appl  
38 34 82.9 1159 3 US-09-358-383C-10 Sequence 10, Appl  
39 34 82.9 1159 4 US-09-275-252A-12 Sequence 12, Appl  
40 34 82.9 1646 4 US-09-252-991A-22312 Sequence 22312, A  
41 34 82.9 1778 4 US-09-252-991A-18159 Sequence 18159, A  
42 34 82.9 1784 3 US-09-040-738-2 Sequence 2, Appl  
43 34 82.9 1784 3 US-08-652-428A-2 Sequence 2, Appl  
44 33 80.5 82 2 US-08-925-237-4 Sequence 4, Appl  
45 33 80.5 109 4 US-09-796-766-16 Sequence 16, Appl

## ALIGNMENTS

### RESULT 1

US-09-489-039A-7834  
; Sequence 7834, Application US/09489039A  
; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 7834

; LENGTH: 423

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-7834

Query Match 92.7%; Score 38; DB 4; Length 423;  
Best Local Similarity 85.7%; Pred. No. 48;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGP 7

Db 194 KGRPQGP 200

### RESULT 2

US-09-252-991A-27110

; Sequence 27110, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 27110

; LENGTH: 446

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27110

```
Query Match      90.2%; Score 37; DB 4; Length 446;
Best Local Similarity 85.7%; Pred. No. 74;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRPQGP 7
Db 405 RGRPRGP 411

RESULT 3
US-09-904-615-157
; Sequence 157, Application US/09904615
; Patent No. 656325
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P203221
; CURRENT APPLICATION NUMBER: US/09/904,615
; CURRENT FILING DATE: 2001-07-16
; PRIOR FILING DATE: 09/511,554
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 157
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-904-615-157

Query Match      87.8%; Score 36; DB 4; Length 58;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGP 7
Db 36 GRPQGP 41

RESULT 4
US-09-370-838-74
; Sequence 74, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-74

Query Match      87.8%; Score 36; DB 4; Length 64;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGP 7
Db 405 RGRPRGP 411
```

```
Db 23 GRPQGP 28

RESULT 5
US-08-355-888A-11
; Sequence 11, Application US/08355888A
; Patent No. 5763211
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/355,888A
; FILING DATE: 14-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 7225-078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 863-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 73 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-355-888A-11

Query Match      87.8%; Score 36; DB 1; Length 73;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGP 7
Db 32 GRPQGP 37

RESULT 6
US-08-693-697-11
; Sequence 11, Application US/08693697
; Patent No. 5869610
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
```

```
/ COUNTRY: USA
/ ZIP: 10036-2711
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: FastSeq for Windows Version 2.0b
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/693,697
/ FILING DATE: 05-AUG-1996
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Poissant, Brian M.
/ REGISTRATION NUMBER: 28,462
/ REFERENCE/DOCKET NUMBER: 8907-0037-999
/ TELEPHONE: 650-493-4935
/ TELEFAX: 650-493-5556
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 11:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 73 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: Protein
/ US-08-693-697-11

Query Match 87.8%; Score 36; DB 2; Length 73;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7
DB 32 GRPQGP 37

RESULT 7
US-08-693-696-11
/ Sequence 11, Application US/08693696
/ Patent No. 6005080
/ GENERAL INFORMATION:
/ APPLICANT: Snodgrass, H. R.
/ APPLICANT: Cioffi, Joseph
/ APPLICANT: Zupancic, Thomas J.
/ APPLICANT: Shafer, Alan W.
/ TITLE OF INVENTION: A NOVEL HUMAN HEMATOPOIETIN
/ NUMBER OF SEQUENCES: 31
/ TITLE OF INVENTION: RECEPTOR
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Pennie & Edmonds
/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10036-2711
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/693,696
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/355,888
/ FILING DATE: 14-DEC-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Poissant, Brian M.
/ REGISTRATION NUMBER: 28,462
/ REFERENCE/DOCKET NUMBER: 7225-078
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 790-9090

/ COUNTRY: USA
/ ZIP: 66141 PENNIE
/ TELEFAX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 11:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 73 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-693-696-11

Query Match 87.8%; Score 36; DB 3; Length 73;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7
DB 32 GRPQGP 37

RESULT 8
US-09-357-914-11
/ Sequence 11, Application US/09357914
/ Patent No. 6524806
/ GENERAL INFORMATION:
/ APPLICANT: Snodgrass, H. Ralph
/ APPLICANT: Cioffi, Joseph
/ APPLICANT: Zupancic, Thomas J.
/ APPLICANT: Shafer, Alan Wayne
/ TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR HU-B1-219, A
/ TITLE OF INVENTION: NOVEL HUMAN HEMATOPOIETIN RECEPTOR
/ FILE REFERENCE: 8907-0083-999
/ CURRENT APPLICATION NUMBER: US/09/357,914
/ CURRENT FILING DATE: 1999-07-19
/ PRIOR APPLICATION NUMBER: US 08/693,696
/ PRIOR FILING DATE: 1996-08-05
/ PRIOR APPLICATION NUMBER: US 08/355,888
/ PRIOR FILING DATE: 1994-12-14
/ PRIOR APPLICATION NUMBER: US 08/306,231
/ PRIOR FILING DATE: 1994-09-14
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 11
/ LENGTH: 73
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-09-357-914-11

Query Match 87.8%; Score 36; DB 4; Length 73;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7
DB 32 GRPQGP 37

RESULT 9
US-09-252-991A-22177
/ Sequence 22177, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 22177
```



```
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-22177

Query Match      87.8%; Score 36; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GRPOGP 7
Db      21 GRPOGP 26

RESULT 10
US-09-252-991A-27546
; Sequence 27546, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27546
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-27546

Query Match      87.8%; Score 36; DB 4; Length 199;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GRPOGP 7
Db      145 GRPOGP 150

RESULT 11
US-09-252-991A-26949
; Sequence 26949, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26949
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-26949

Query Match      87.8%; Score 36; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GRPOGP 7
Db      145 GRPOGP 150
```

```
Db      19 GRPOGP 24

RESULT 12
US-09-252-991A-16906
; Sequence 16906, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16906
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-16906

Query Match      87.8%; Score 36; DB 4; Length 263;
Best Local Similarity 85.7%; Pred. No. 66;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 RGRPOGP 7
Db      90 RGRPOGP 96

RESULT 13
US-09-252-991A-21690
; Sequence 21690, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21690
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-21690

Query Match      87.8%; Score 36; DB 4; Length 473;
Best Local Similarity 85.7%; Pred. No. 11e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 RGRPOGP 7
Db      317 RGRPOGP 323

RESULT 14
US-08-640-389A-8
; Sequence 8, Application US/08640389A
; Patent No. 5912123
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
```

APPLICANT: Shafer, Alan W.  
TITLE OF INVENTION: DETECTION OF THE LEFTIN  
TITLE OF INVENTION: RECEPTOR IN REPRODUCTIVE ORGANS AND METHODS FOR  
TITLE OF INVENTION: REGULATING REPRODUCTIVE BIOLOGY  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/640,389A  
FILING DATE: 29-APR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Poissant, Brian M.  
REGISTRATION NUMBER: 28,462  
REFERENCE/DOCKET NUMBER: 8907-032  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 958 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-640-389A-8

Query Match 87.8%; Score 36; DB 2; Length 958;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGP 7  
Db 917 GRPQGP 922

RESULT 15  
US-08-618-957A-8  
Sequence 8, Application US/08618957A  
Patent No. 6355237  
GENERAL INFORMATION:  
APPLICANT: Snodgrass, H. Ralph  
APPLICANT: Cioffi, Joseph  
APPLICANT: Zupancic, Thomas Joel  
APPLICANT: Shafer, Alan Wayne  
TITLE OF INVENTION: METHODS FOR USING THE OBSE  
TITLE OF INVENTION: GENE AND ITS GENE PRODUCT TO STIMULATE HEMATOPOIETIC  
TITLE OF INVENTION: DEVELOPMENT  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of The Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2811  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/618,957A  
FILING DATE: 20-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Poissant, Brian M.  
REGISTRATION NUMBER: 28,462  
REFERENCE/DOCKET NUMBER: 008907-0033-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-493-4935  
TELEFAX: 650-493-5556  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 958 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-618-957A-8  
Query Match 87.8%; Score 36; DB 4; Length 958;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GRPQGP 7  
Db 917 GRPQGP 922  
Search completed: April 6, 2004, 16:19:39  
Job time: 11.271 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 8.97196 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-4

Perfect score: 48

Sequence: 1 RGRPQGGP 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	166	1 PIHUSC	salivary proline-r
2	48	100.0	166	2 B25372	salivary proline-r
3	48	100.0	171	2 A27307	proline-rich phosp
4	43	89.6	212	2 B36298	proline-rich prote
5	40	83.3	176	2 C38355	basic proline-rich
6	40	83.3	117	2 D40750	proline-rich prote
7	40	83.3	128	2 D38355	basic proline-rich
8	40	83.3	147	2 A29149	proline-rich prote
9	40	83.3	164	2 A30496	proline-rich prote
10	40	83.3	170	2 A48013	proline-rich prote
11	40	83.3	172	2 B29149	proline-rich prote
12	40	83.3	188	2 JH0481	basic proline-rich
13	40	83.3	204	2 A39066	proline-rich prote
14	40	83.3	206	1 PIR73	acidic proline-ric
15	40	83.3	223	2 A42817	proline-rich prote
16	40	83.3	251	1 PIHUPP	salivary proline-r
17	40	83.3	309	2 S10889	proline-rich prote
18	40	83.3	310	1 PIHUSD	salivary proline-r
19	40	83.3	392	1 PIHUB6	salivary proline-r
20	39	81.2	697	2 T00267	hypothetical prote
21	39	81.2	1736	2 T00391	hypothetical prote
22	38	79.2	11	1 XAVIEB	bradykinin-potenti
23	38	79.2	188	2 D29149	proline-rich prote
24	38	79.2	240	2 B24264	proline-rich prote
25	38	79.2	295	2 B48013	proline-rich prote
26	38	79.2	317	2 A28956	proline-rich prote
27	38	79.2	416	2 S52078	proline-rich prote
28	38	79.2	811	2 T36581	probable transmemb
29	37	77.1	220	2 A36298	proline-rich prote

## ALIGNMENTS

### RESULT 1

#### PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N/Alternate names: salivary acidic proline-rich protein PRH2

C/Species: Homo sapiens (man)

C/Date: 31-Mar-1981 #sequence revision 12-Apr-1996 #text change 08-Dec-2000

C/Accession: A25372; A19803; B57868; A92277; A92254; A34425; A91954; S02564; S02563; JP

R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A/Title: Differential RNA splicing and post-translational cleavages in the human saliva

A/Reference number: A92492; MUID:85289325; PMID:2993301

A/Accession: A25372

A/Molecule type: mRNA

A/Residues: 1-166 <KIM>

A/Cross-references: GB:K03202; NID:g190481; PIDN:AAA60183.1; PID:g190482

R/Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A/Title: Primary structure of the active tryptic fragments of human and monkey salivary

A/Reference number: A91757; MUID:81191179; PMID:7228490

A/Accession: A19803

A/Molecule type: protein

A/Residues: 17-46 <SCH>

R/Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A/Title: Structures of two HaeIII-type genes in the human salivary proline-rich protein

A/Reference number: A57868; MUID:86196106; PMID:3009472

A/Accession: B57868

A/Molecule type: DNA

A/Residues: 1-166 <KIM>

A/Cross-references: GB:M13058; NID:g190513; PIDN:AAA98908.1; PID:g190514

R/Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A/Title: The primary structure of a salivary calcium-binding proline-rich phosphoprotein:

A/Reference number: A92277; MUID:80204368; PMID:7380845

A/Contents: protein C

A/Accession: A92277

A/Molecule type: protein

A/Residues: 17-19,'N',21-166 <WON>

A/Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite form.

R/Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A/Title: The complete primary structure of a proline-rich phosphoprotein from human sal

A/Reference number: A92254; MUID:79173237; PMID:438215

A/Contents: protein A

A/Accession: A92254

A/Molecule type: protein

A/Residues: 17-19,'N',21-122 <WO2>

R/Schlesinger, D.H.; Hay, D.I.

A/Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A/Reference number: A94425

RESULT 2  
B25372

C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 29-Aug-1997  
C;Accession: A27307  
R;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
Am. J. Hum. Genet. 41, 1035-1047, 1987  
A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
A;Reference number: A27307; MUID:88074309; PMID:3687941  
A;Accession: A27307  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-171 <AZE>  
A;Cross-references: EMBL:K03203  
C;Gene: GDB:PRH1  
A;Cross-references: GDB:119515; OMIM:168730  
A;Map position: 12p13.2-12p13.2  
C;Superfamily: proline-rich protein  
C;Keywords: phosphoprotein

Query Match 100.0%; Score 48; DB 2; Length 171;  
Best Local Similarity 100.0%; Pred. No. 0.68; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGGP 8  
| | | | |  
DB 127 RGRPQGGP 134

RESULT 4  
B36298  
proline-rich protein PRB3S (cys) - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 29-Aug-1997  
C;Accession: B36298  
R;Azen, E.A.; Minaguchi, K.; Latreille, P.; Kim, H.S.  
Am. J. Hum. Genet. 47, 686-697, 1990  
A;Title: Alleles at the PRB3 locus coding for a disulfide-bonded human salivary proline-rich protein  
A;Reference number: A36298; MUID:91022705; PMID:2171329  
A;Accession: B36298  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-212 <AZE>  
C;Genetics:  
A;Gene: GDB:PRB3  
A;Cross-references: GDB:119513; OMIM:168840  
A;Map position: 12p13.2-12p13.2  
C;Superfamily: proline-rich protein

Query Match 89.6%; Score 43; DB 2; Length 212;  
Best Local Similarity 100.0%; Pred. No. 5.4; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
| | | | |  
DB 167 GRPQGGP 173

RESULT 5  
C38355  
basic proline-rich peptide II-2 - human  
C;Species: Homo sapiens (man)  
C;Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 12-Apr-1995  
C;Accession: C38355  
R;Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
Biochemistry 30, 3351-3356, 1991  
A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov  
A;Reference number: A38355; MUID:91190884; PMID:1849422  
A;Accession: C38355  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-76 <KAU>  
C;Superfamily: proline-rich protein

Query Match 83.3%; Score 40; DB 2; Length 76;  
Best Local Similarity 85.7%; Pred. No. 11; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
| | | | |  
DB 4 GRPQGGP 10

RESULT 8  
A29149  
proline-rich protein - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 20-Aug-1999  
C;Accession: A29149  
R;Clements, S.; Mehansho, H.; Carlson, D.M.  
J. Biol. Chem. 260, 13471-13477, 1985  
A;Title: Novel multigene families encoding highly repetitive peptide sequences. Sequenc  
A;Reference number: A92501; MUID:86033799; PMID:3840480  
A;Accession: A29149  
A;Molecule type: mRNA  
A;Residues: 1-147 <CLE>  
A;Cross-references: GB:M11900; NID:g200538; PIDN:AAA40000.1; PID:g200539

Best Local Similarity 85.7%; Pred. No. 6.5; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
| | | | |  
DB 40 GRPQGGP 46

RESULT 6  
D40750  
proline-rich protein PRB1/2S (EA) - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 03-May-1996  
C;Accession: D40750  
R;Azen, E.A.; Latreille, P.; Niece, R.L.  
Am. J. Hum. Genet. 53, 264-278, 1993  
A;Title: PRB1 gene variants coding for length and null polymorphisms among human saliva  
A;Reference number: A40750; MUID:93304421; PMID:8317492  
A;Accession: D40750  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-117 <AZE>  
A;Cross-references: GB:S62930  
C;Superfamily: proline-rich protein

Query Match 83.3%; Score 40; DB 2; Length 117;  
Best Local Similarity 85.7%; Pred. No. 9.7; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
| | | | |  
DB 21 GRPQGGP 27

RESULT 7  
D38355  
basic proline-rich peptide IB-8a - human (fragments)  
C;Species: Homo sapiens (man)  
C;Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 12-Apr-1995  
C;Accession: D38355  
R;Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
Biochemistry 30, 3351-3356, 1991  
A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the co  
A;Reference number: A38355; MUID:91190884; PMID:1849422  
A;Accession: D38355  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-128 <KAU>  
C;Superfamily: proline-rich protein

Query Match 83.3%; Score 40; DB 2; Length 128;  
Best Local Similarity 85.7%; Pred. No. 11; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
| | | | |  
DB 4 GRPQGGP 10



C;Superfamily: proline-rich protein

Query Match 83.3%; Score 40; DB 2; Length 204;  
Best Local Similarity 85.7%; Pred. No. 16;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
|:|||||  
Db 104 GKPQGGP 110

RESULT 14

PIRT3  
acidic proline-rich protein precursor - rat  
N;Alternate names: PRP  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 22-Jun-1999  
C;Accession: A03296  
R;Ziemer, M.A.; Swain, W.F.; Rutter, W.J.; Clements, S.; Ann, D.K.; Carlson, D.M.  
J. Biol. Chem. 259, 10475-10480, 1984  
A;Title: Nucleotide sequence analysis of a proline-rich protein cDNA and peptide homolog  
A;Reference number: A03296; MUID:84289443; PMID:6547951  
A;Accession: A03296  
A;Molecule type: mRNA  
A;Residues: 1-206 <ZIE>  
A;Cross-references: GB:K02247; NID:G206395; PIDN:AAA1949.1; PID:G206396  
C;Comment: This protein contains six 18- to 19-residue repeats.  
C;Comment: This protein may protect teeth by binding to tannins.  
C;Superfamily: proline-rich protein  
C;Keywords: duplication; parotid gland; saliva; tandem repeat  
F;1-13/Domain: signal sequence #status predicted <SIG>  
F;14-206/Product: acidic proline-rich protein #status predicted <MAT>  
F;80-189/Region: 18-residue repeats

Query Match 83.3%; Score 40; DB 1; Length 206;  
Best Local Similarity 85.7%; Pred. No. 16;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
|:|||||  
Db 148 GKPQGGP 154

RESULT 15

A42817  
proline-rich protein precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 26-May-1994 #sequence\_revision 01-Dec-1995 #text\_change 17-Mar-1999  
C;Accession: A42817  
R;Castle, A.M.; Stahl, L.E.; Castle, J.D.  
J. Biol. Chem. 267, 13093-13100, 1992  
A;Title: A 13-amino acid N-terminal domain of a basic proline-rich protein is necessary  
A;Reference number: A42817; MUID:92317007; PMID:1618808  
A;Accession: A42817  
A;Molecule type: mRNA  
A;Residues: 1-223 <CAS>  
A;Cross-references: GB:M83567  
A;Note: nucleotide sequence not given; translation not complete  
C;Superfamily: proline-rich protein  
C;Keywords: glycoprotein; saliva  
F;1-16/Domain: signal sequence #status predicted <SIG>

Query Match 83.3%; Score 40; DB 2; Length 223;  
Best Local Similarity 85.7%; Pred. No. 17;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
|:|||||  
Db 85 GKPQGGP 91

Search completed: April 6, 2004, 16:16:51  
Job time : 9.97196 secs

GenCore version 5.1.6.  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 5.15888 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-4

Perfect score: 48

Sequence: 1 RGRPGGP 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	166	1 PRPC_HUMAN	P02810 homo sapien
2	42	87.5	124	1 CSR4_HUMAN	Q9bxc7 homo sapien
3	40	83.3	61	1 PRP5_HUMAN	P02811 homo sapien
4	40	83.3	96	1 PRP2_RAT	P04281 homo sapien
5	40	83.3	172	1 PRP2_HUMAN	P10164 rattus norv
6	40	83.3	174	1 PRPP_HUMAN	P81489 homo sapien
7	40	83.3	206	1 PRP3_RAT	P04474 rattus norv
8	40	83.3	234	1 PRPM_HUMAN	P10161 homo sapien
9	40	83.3	247	1 PRP4_HUMAN	P10163 homo sapien
10	40	83.3	251	1 PRP2_HUMAN	P02812 homo sapien
11	40	83.3	276	1 PRP1_HUMAN	P10162 homo sapien
12	40	83.3	331	1 PRP1_HUMAN	P04280 homo sapien
13	38	79.2	11	1 BPP_AKGHP	P04562 agkistrodon
14	38	79.2	296	1 PMB3_MOUSE	P05143 mus musculu
15	38	79.2	416	1 P12R_RAT	P43253 rattus norv
16	38	79.2	842	1 ORP7_HUMAN	Q92zf2 homo sapien
17	38	79.2	5065	1 EPPL_HUMAN	P88107 homo sapien
18	37	77.1	113	1 VAM2_XENLA	P47193 xenopus lae
19	37	77.1	261	1 PRP2_MOUSE	P05142 mus musculu
20	37	77.1	323	1 LEG4_HUMAN	P56470 homo sapien
21	37	77.1	740	1 GAG_SVRVH	P14111 squirrel mo
22	37	77.1	797	1 PAT1_YEAST	P25644 saccharomyc
23	37	77.1	1806	1 CALB_HUMAN	P12107 homo sapien
24	37	77.1	5262	1 MLL2_HUMAN	O44686 homo sapien
25	36	75.0	125	1 SYB_LOLPE	P47194 loligo peal
26	36	75.0	209	1 NT5_RAT	P34131 rattus norv
27	36	75.0	210	1 NT5_HUMAN	P34130 homo sapien
28	36	75.0	393	1 B4G3_HUMAN	O60512 h.beta-1.4-
29	36	75.0	520	1 GAG_SIVAL	P27972 simian immu
30	36	75.0	548	1 ERF_HUMAN	P50548 homo sapien
31	36	75.0	551	1 ERF_MOUSE	P70459 mus musculu
32	36	75.0	740	1 CEB1_HUMAN	Q15027 homo sapien
33	36	75.0	1226	1 CDN1_HUMAN	Q81wy9 homo sapien

RESULT 1  
PRPC\_HUMAN  
ID PRPC\_HUMAN STANDARD; PRT; 166 AA.  
AC P02810;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains: Peptide P-C].  
DE Peptide P-C.  
GN PRH1 AND PRH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).  
RX MEDLINE=86196106; PubMed=3009472;  
RA Kim H.-S., Maeda N.;  
RT "Structures of two HaeIII-type genes in the human salivary proline-rich protein multigene family";  
RL J. Biol. Chem. 261:6712-6718(1986).  
RN [2]  
RP SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).  
RX MEDLINE=85289325; PubMed=2993301;  
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system";  
RL J. Biol. Chem. 260:11123-11130(1985).  
RN [3]  
RP SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).  
RX MEDLINE=89061650; PubMed=3196309;  
RA Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K., Madapallinattam G., Schluckebier S.K.;  
RT "The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";  
RL Biochem. J. 255:15-21(1988).  
RN [4]  
RP SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).  
RX MEDLINE=88074309; PubMed=3687941;  
RA Azen E.A., Kim H.S., Goodman P., Flynn S., Maeda N.;  
RT "Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich proteins Pa, Db, and Pif.";  
RL Am. J. Hum. Genet. 41:1035-1047(1987).  
RN [5]  
RP SEQUENCE OF 17-166 (PRP-2).  
RX MEDLINE=86222916; PubMed=3710693;  
RA Schlesinger D.H., Hay D.I.;  
RT "Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva.";  
RL Int. J. Pept. Protein Res. 27:373-379(1986).  
RN [6]  
RP SEQUENCE OF 17-166 (PROTEIN C).  
RX MEDLINE=80204368; PubMed=7380845;  
RA Wong R.S.C., Bennick A.;

P97260 cricetus  
Q12770 homo sapien  
Q9npd0 homo sapien  
C8f1l3 corynebacte  
F03291 human adeno  
Q9kv09 vibrio choi  
P40605 vibrio para  
P25662 escherichia  
P21918 homo sapien  
Q43390 homo sapien  
Q04691 drosophila  
P41110 oryctolagus



RT "The primary structure of a salivary calcium-binding proline-rich  
RT phosphoprotein (protein C), a possible precursor of a related  
RT salivary protein A.";  
RL J. Biol. Chem. 255:5943-5948(1980).  
RN [7]  
RP SEQUENCE OF 17-46 (PROTEIN C).  
RX MEDLINE=8191179; PubMed=7228490;  
RA Schlesinger D.H., Hay D.I.;  
RT "Primary structure of the active tryptic fragments of human and  
RT monkey salivary anionic proline-rich proteins.";  
RL Int. J. Pept. Protein Res. 17:34-41(1981).  
RN [8]  
RP SEQUENCE OF 17-122 (PROTEIN A).  
RX MEDLINE=79173237; PubMed=438215;  
RA Wong R.S.C., Hofmann T., Bennick A.;  
RT "The complete primary structure of a proline-rich phosphoprotein from  
RT human saliva.";  
RL J. Biol. Chem. 254:4800-4808(1979).  
RN [9]  
RP SEQUENCE OF 17-122 (PROTEIN A).  
RA Schlesinger D.H., Hay D.I.;  
RT "Complete primary structure of a proline-rich phosphoprotein (PRP-4),  
RT a potent inhibitor of calcium phosphate precipitation in human parotid  
RT saliva.";  
RL (In) Gross E., Meienhofer J. (eds.);  
RL Peptides: structure and biological function (Proceedings of the 6th  
RL American peptide symposium), pp.133-136, Pierce Chemical Co.,  
RL Rockford IL. (1979).  
RN [10]  
RP SEQUENCE OF 123-166 (PEPTIDE P-C).  
RX MEDLINE=80227634; PubMed=7390979;  
RA Isemura S., Saitoh E., Sanada K.;  
RT "The amino acid sequence of a salivary proline-rich peptide, P-C, and  
RT its relation to a salivary proline-rich phosphoprotein, protein C.";  
RL J. Biochem. 87:1071-1077(1980).  
RN [11]  
RP VARIANT PRH2-3 LYS-163.  
RA Azen E.A.;  
RT "A frequent mutation in the acidic proline-rich protein gene, PRH2,  
RT causing a Q14YK change closely adjacent to the bacterial binding  
RT domain of the cognate salivary PRP (Pr1) in Afro-Americans.";  
RL Hum. Mutat. 12:72-72(1998).  
CC -1- FUNCTION: PRP's act as highly potent inhibitors of crystal growth  
CC of calcium phosphates. They provide a protective and reparative  
CC environment for dental enamel which is important for the integrity  
CC of the teeth.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIF-S yield PRP-4,  
CC PRP-3 (protein A), and PIF-P, respectively.  
CC -1- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele;  
CC allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is  
CC also known as PR1'.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; K03202; AAA60183.1; -  
DR EMBL; K03203; AAA60184.1; -  
DR EMBL; M13057; AAA98807.1; -  
DR EMBL; M13058; AAA98808.1; -  
DR Genbank; HGNC:9366; PRH1.  
DR Genbank; HGNC:9367; PRH2.  
DR MIM; 168730; -  
DR MIM; 168790; -  
DR MIM; 168710; -  
DR GO; GO:0005615; Cytoplasmic space; TAS.  
DR Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism;  
KW Pyrrolidone carboxylic acid.

FT SIGNAL 1 16  
FT CHAIN 17 166  
FT CHAIN 17 122  
FT CHAIN 123 166  
FT DOMAIN 17 46  
FT MOD\_RES 17 17  
FT MOD\_RES 24 24  
FT MOD\_RES 38 38  
FT VARIANT 20 20  
FT VARIANT 66 66  
FT VARIANT 163 163  
FT CONFLICT 41 41  
FT SEQUENCE 166 AA; 17017 MW; A7DF62BF94E33EF CRC64;  
SQ  
Query Match 100.0%; Score 48; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 0.34; Mismatches 0; Gaps 0;  
Matches 8; Conservative 0; Indels 0;  
QY 1 RGRPQGGP 8  
DB 122 RGRPQGGP 129  
RESULT 2  
ID CES4 HUMAN STANDARD; PR1; 124 AA.  
AC Q9BXQ7;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cat eye syndrome critical region protein 4 (Fragment).  
GN CECR4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21275466; PubMed=11381032;  
RA Footz T.K., Brinkman-Mills P., Banning G.S., Maier S.A., Riazzi M.A.,  
RA Bridgland L.J., Hu S., Bixren B., Minoshima S., Shimizu N., Fan H.,  
RA Nguyen T., Fang F., Fu Y., Ray L., Wu H., Shaull S., Phan S., Yao Z.,  
RA Chen F., Huan A., Hu P., Wang Q., Loh P., Qi S., Roe B.A.,  
RA McDermid H.E.;  
RT "Analysis of the cat eye syndrome critical region in humans and the  
RT region of conserved synteny in mice: a search for candidate genes at  
RT or near the human chromosome 22 pericentromere.";  
RL Genome Res. 11:1053-1070(2001).  
CC -1- TISSUE SPECIFICITY: Adult heart and skeletal muscle. Widely  
CC expressed in fetal tissues.  
CC -1- MISCELLANEOUS: Candidate gene for the Cat Eye Syndrome (CES), a  
CC developmental disorder associated with the duplication of a 2 Mb  
CC region of 22q11.2. Duplication usually takes in the form of a  
CC supernumerary bisatellited isocentric chromosome, resulting in  
CC four copies of the region (represents an inv dup(22)(q11)). CES is  
CC characterized clinically by the combination of coloboma of the  
CC iris and anal atresia with fistula, downslanting palpebral  
CC fissures, preauricular tags and/or pits, frequent occurrence of  
CC heart and renal malformations, and normal or near-normal mental  
CC development.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF307448; AAK30048.1; -.
CC Genew; HGNC:1842; CECR4.
DR FT NON_TER 1
SQ SEQUENCE 124 AA; 12838 MW; 938E00386308EC7A CRC64;
Query Match 87.5%; Score 42; DB 1; Length 124;
Best Local Similarity 87.5%; Pred. No. 2.4;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 RGRPGQPP 8
DB 58 RGRPOAPP 65
RESULT 3
PRPE_HUMAN
ID PRPE_HUMAN STANDARD; PRT; 61 AA.
AC P02811;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Basic proline-rich peptide P-E (IB-9).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE
RC TISSUE=Saliva;
RX MEDLINE=83101329; PubMed=6924859;
RA Kaufman D., Wong R., Bennick A., Keller P.;
RT "Basic proline-rich proteins from human parotid saliva: complete
RT covalent structure of protein IB-9 and partial structure of protein
RT IB-6, members of a polymorphic pair.";
RL Biochemistry 21:6558-6562(1982).
RN [2]
RP SEQUENCE
RC TISSUE=Saliva;
RX MEDLINE=83007119; PubMed=7118863;
RA Isemura S., Saitoh E., Sanada K.;
RT "Fractionation and characterization of basic proline-rich peptides of
RT human parotid saliva and the amino acid sequence of proline-rich
RT peptide P-E.";
RL J. Biochem. 91:2067-2075(1982).
CC -!- MISCELLANEOUS: Peptides IB-9 and P-E are the same peptide.
DR GO; GO:0005576; C:extracellular; NAS.
KW Repeat; Parotid gland.
FT REPEAT 1 21 APPROXIMATE.
FT REPEAT 22 42 APPROXIMATE.
FT REPEAT 43 61 APPROXIMATE.
SQ SEQUENCE 61 AA; 6024 MW; F883F78617B05327 CRC64;
Query Match 83.3%; Score 40; DB 1; Length 61;
Best Local Similarity 85.7%; Pred. No. 2.5;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 GRPGQPP 8
DB 4 GRPQGGP 10
RESULT 4
PRP5_HUMAN
ID PRP5_HUMAN STANDARD; PRT; 96 AA.
AC P04281;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Basic proline-rich peptide IB-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE
RC TISSUE=Saliva;
RX MEDLINE=86243355; PubMed=3521730;
RA Kauffman D., Hofmann T., Bennick A., Keller P.;
RT "Basic proline-rich proteins from human parotid saliva: complete
RT covalent structures of proteins IB-1 and IB-6.";
RL Biochemistry 25:2387-2392(1986).
CC -!- DOMAIN: Contains 21-residue repeats, two of which have internal
CC 7-residue repeats.
CC -!- MISCELLANEOUS: There are nine basic proline-rich peptides isolated
CC from the saliva; this peptide is designated IB-1.
DR PIR; B40750; PIHU6.
KW Phosphorylation; Repeat; Parotid gland; Pyrrolidone carboxylic acid.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 8 8 PHOSPHORYLATION.
FT REPEAT 15 35 APPROXIMATE.
FT REPEAT 36 56 APPROXIMATE.
FT REPEAT 57 77 APPROXIMATE.
SQ SEQUENCE 96 AA; 9530 MW; FF3B33F68DE104F8 CRC64;
Query Match 83.3%; Score 40; DB 1; Length 96;
Best Local Similarity 85.7%; Pred. No. 3.9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 GRPGQPP 8
DB 39 GRPQGGP 45
RESULT 5
PRP2_RAT
ID PRP2_RAT STANDARD; PRT; 172 AA.
AC P10164;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acidic proline-rich protein PRP25 precursor (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Parotid gland;
RX MEDLINE=86033799; PubMed=3840480;
RA Clements S., Mehansho H., Carlson D.M.;
RT "Novel multigene families encoding highly repetitive peptide
RT sequences. Sequence analyses of rat and mouse proline-rich protein
RT cDNAs.";
RL J. Biol. Chem. 260:13471-13477(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M11901; AAA41948.1; -.
DR PIR; B29149; B29149.
KW Repeat; Parotid gland; Multigene family; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 >172 ACIDIC PROLINE-RICH PROTEIN PRP25.
FT NON_TER 172 172
SQ SEQUENCE 172 AA; 17416 MW; F63BFBD05459D6EA CRC64;
Query Match 83.3%; Score 40; DB 1; Length 172;
Best Local Similarity 85.7%; Pred. No. 6.9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; K02247; AAA1949.1; -;  
DR EMBL; M11898; AAA1958.1; -;  
DR PIR; A03296; PIRT3.  
KW Repeat; Parotid gland; Multigene family; Signal.  
FT SIGNAL 1 13 POTENTIAL.  
FT CHAIN 14 206 ACIDIC PROLINE-RICH PROTEIN PRP33.  
FT DOMAIN 80 189 6 X 18 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 80 97 1.  
FT REPEAT 98 115 2.  
FT REPEAT 116 133 3.  
FT REPEAT 134 152 4.  
FT REPEAT 153 170 5.  
FT REPEAT 171 189 6.  
SQ SEQUENCE 206 AA; 21403 MW; D037582CEDD05CE6B CRC64;  
  
Query Match 83.3%; Score 40; DB 1; Length 206;  
Best Local Similarity 85.7%; Pred. No. 8.3; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0;  
  
QY 2 GRPOGPP 8  
|:|||||  
Db 148 GKPOGPP 154  
  
RESULT 8  
PRPM\_HUMAN  
ID PRPM\_HUMAN STANDARD; PRT; 234 AA.  
AC P10161; P02813;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary proline-rich protein PO (Allele M) [Contains: Peptide P-D]  
DE (Fragment).  
DE PR84.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=89121439; PubMed=3220251;  
RA Lyons K.M., Stein J.H., Smithies O.;  
RT "Many protein products from a few loci: assignment of human salivary  
RT proline-rich proteins to specific loci."  
RL Genetics 120:255-265(1988).  
RN [2]  
RP SEQUENCE OF 165-234.  
RX MEDLINE=83186122; PubMed=6841349;  
RA Saitoh E., Isemura S., Sanada K.;  
RT "Complete amino acid sequence of a basic proline-rich peptide, P-D,  
RT from human parotid saliva."  
RJ J. Biochem. 93:495-502(1983).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/annotation/](http://www.isb-sib.ch/announcement/))  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [license@isb-sib.ch](http://www.isb-sib.ch/announcement/</a>)<br/>CC or send an email to <a href=).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [license@isb-sib.ch](http://www.isb-sib.ch/announcement/</a>)<br/>CC or send an email to <a href=).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [license@isb-sib.ch](http://www.isb-sib.ch/announcement/</a>)<br/>CC or send an email to <a href=).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [license@isb-sib.ch](http://www.isb-sib.ch/announcement/</a>)<br/>CC or send an email to <a href=).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [license@isb-sib.ch](http://www.isb-sib.ch/announcement/</a>)<br/>CC or send an email to <a href=).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [license@isb-sib.ch](http://www.isb-sib.ch/announcement/</a>)<br/>CC or send an email to <a href=).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [license@isb-sib.ch](http://www.isb-sib.ch/announcement/</a>)<br/>CC or send an email to <a href=).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [license@isb-sib.ch](http://www.isb-sib.ch/announcement/</a>)<br/>CC or send an email to <a href=).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [license@isb-sib.ch](http://www.isb-sib.ch/announcement/</a>)<br/>CC or send an email to <a href=).  
CC -----  
DR EMBL; X07704; CAA30542.1; -;  
DR MIM; 168730; -;  
DR MIM; 180990; -;

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified

```
DR GO: 0005576; C:extracellular; NAS
KW Repeat; Parotid gland; Multigene family.
FT NON_TER 1
FT CHAIN 165 234 PEPTIDE P-D.
SQ SEQUENCE 234 AA; 23676 MW; 310AFF13A44E747F CRC64;

Query Match 83.3%; Score 40; DB 1; Length 234;
Best Local Similarity 85.7%; Pred. No. 9.5;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8
Db 21 GKPGQGP 27

RESULT 9
PRP4_HUMAN STANDARD; PRT; 247 AA.
AC P10163; P02813;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein precursor (Allele S).
GN PRB4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85289325; PubMed=2993301;
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT "Differential RNA splicing and post-translational cleavages in the
human salivary proline-rich protein gene system.";
RL J. Biol. Chem. 260:11123-11130(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89121440; PubMed=2851479;
RA Lyons K.M., Stein J.H., Smithies O.;
RT "Length polymorphisms in human proline-rich protein genes generated
by intragenic unequal crossing over.";
RL Genetics 120:267-278(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03207; AAA60188.1; -.
DR EMBL; X07882; CAA30729.1; -.
DR PIR; S01176; PIHUSD.
DR Genew; HGNC:9340; PRB4.
DR MIM; 168730; -.
DR MIM; 180990; -.
KW Repeat; Parotid gland; Multigene family; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 247 SALIVARY PROLINE-RICH PROTEIN PO.
FT CHAIN 17 39 PROTEIN N1.
FT CHAIN 40 177 GLYCOSYLATED PROTEIN A.
FT CHAIN 37 57 MISSING (IN REF. 2).
FT CONFLICT 218 218 D -> A (IN REF. 2).
FT CONFLICT 218 218 D -> A (IN REF. 2).
SQ SEQUENCE 247 AA; 25108 MW; 6A1943E435161691 CRC64;

Query Match 83.3%; Score 40; DB 1; Length 247;
Best Local Similarity 85.7%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8
Db 55 GKPGQGP 61

RESULT 10
PRP2_HUMAN STANDARD; PRT; 251 AA.
AC P02812;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein precursor (Clone CP7) [Contains: Basic
peptide P-F] (fragment).
GN PRB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85289325; PubMed=2993301;
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT "Differential RNA splicing and post-translational cleavages in the
human salivary proline-rich protein gene system.";
RL J. Biol. Chem. 260:11123-11130(1985).
RN [2]
RP SEQUENCE OF 134-194.
RX MEDLINE=83265674; PubMed=6874669;
RA Saitoh E., Isemura S., Sanada K.;
RT "Complete amino acid sequence of a basic proline-rich peptide, P-F,
from human parotid saliva.";
RL J. Biochem. 93:883-888(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03208; AAA60189.1; -.
DR PIR; E25372; PIHUPF.
DR Genew; HGNC:9338; PRB2.
DR MIM; 168810; -.
KW Repeat; Parotid gland; Multigene family.
FT NON_TER 1 1
FT CHAIN 134 194 BASIC PEPTIDE P-F.
FT REPEAT <1 9
FT REPEAT 10 71
FT REPEAT 72 133
FT REPEAT 134 195
SQ SEQUENCE 251 AA; 24641 MW; D779F590C0EBF30B CRC64;

Query Match 83.3%; Score 40; DB 1; Length 251;
Best Local Similarity 85.7%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8
Db 13 GKPGQGP 19

RESULT 11
PRPL_HUMAN STANDARD; PRT; 276 AA.
AC P10162; P02813;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein PO (Allele X) [Contains: Peptide P-D]
(fragment).
GN PRB4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89121439; PubMed=3220251;
RA Lyons K.M., Stein J.H., Smithies O.;
RT "Many protein products from a few loci: assignment of human salivary
RL proline-rich proteins to specific loci.";
RL Genetics 120:255-265(1988).
RN [2]
RP SEQUENCE OF 207-276
RX MEDLINE=83185122; PubMed=6841349;
RA Saitoh E., Isemura S., Sanada K.;
RT "Complete amino acid sequence of a basic proline-rich peptide, P-D,
RL from human parotid saliva.";
RL J. Biochem. 93:495-502(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07715; CAA30543.1; ALT_SEQ.
DR F1R; S03176; PIHUSD.
DR MIM; 168730; -.
DR MIM; 180990; -.
DR GO; GO:0005576; C:extracellular; NAS.
DR INTERPRO; IPR00637; AT hook.
DR PRINTS; PR00929; ATHOOK.
KW Repeat; Parotid gland; Multigene family.
FT NON_TER 1 276 PEPTIDE P-D.
FT CHAIN 207 276
FT SEQUENCE 276 AA; 27816 MW; 9F494926C979441A CRC64;
Query Match 83.3%; Score 40; DB 1; Length 276;
Best Local Similarity 85.7%; Pred. No. 11;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 GRPQGP 8
Db 21 GKPGGP 27
RESULT 12
ID PRP1_HUMAN STANDARD; PRT; 331 AA.
AC P04280;
DT 20-MAR-1987 (Rel. 04, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein precursor (Clones CP3, CP4 and CP5)
DE (Contains: Basic peptide IB-6; Peptide P-H).
GN PRB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85289325; PubMed=2993301;
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT "Differential RNA splicing and post-translational cleavages in the
RL human salivary proline-rich protein gene system.";
RL J. Biol. Chem. 260:11123-11130(1985).
RN [2]
RP SEQUENCE OF 214-331.
RX MEDLINE=86243355; PubMed=3521730;
RA Kauffman D., Hofmann T., Bennick A., Keller P.;
RT "Basic proline-rich proteins from human parotid saliva: complete
RL covalent structures of proteins IB-1 and IB-6.";

```

```

RL Biochemistry 25:2387-2392(1986).
RN [3]
RP SEQUENCE OF 276-331.
RX MEDLINE=84161824; PubMed=6671974;
RA Saitoh E., Isemura S., Sanada K.;
RT "Further fractionation of basic proline-rich peptides from human
RL parotid saliva and complete amino acid sequence of basic proline-rich
RL peptide P-H.";
RL J. Biochem. 94:1991-1997(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03204; AAA60185.1; -.
DR EMBL; K03205; AAA60186.1; -.
DR EMBL; K03206; AAA60187.1; -.
DR PIR; B40750; PIHUB6
DR Genew; HGNC:19337; PRB1.
DR MIM; 180989; -.
DR Repeat; Parotid gland; Multigene family; Signal.
FT SIGNAL 1 16
FT CHAIN 214 331 PEPTIDE IB-6.
FT CHAIN 276 331 PEPTIDE P-H.
FT VARIANT 106 238 Missing (in clone CP-4).
FT VARIANT 106 258 /FTIDVAR 005561.
FT VARIANT 276 276 Missing (in clone CP-5).
FT CONFLICT 276 276 A -> S (IN REF. 2 AND 3).
FT SEQUENCE 331 AA; 32596 MW; 3F481FF8EBA39751 CRC64;
Query Match 83.3%; Score 40; DB 1; Length 331;
Best Local Similarity 85.7%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 GRPQGP 8
Db 55 GKPGGP 61
RESULT 13
ID BPP AGKHP STANDARD; PRT; 11 AA.
AC P04562;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bradykinin-potentiating peptide (Angiotensin-converting
DE enzyme inhibitor).
OS Agkistrodon halys pallas (Chinese water moccasin) (Gloydius halys
OS pallas).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Gloydius.
OX NCBI_TaxID=8714;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=86177022; PubMed=3008123;
RA Chi C.-W., Wang S.-Z., Xu L.-G., Wang M.-Y., Lo S.-S., Huang W.-D.;
RT "Structure-function studies on the bradykinin potentiating peptide
RL from Chinese snake venom (Agkistrodon halys pallas)".;
RL Peptides 6 Suppl. 3:339-342(1985).
CC -!- FUNCTION: This peptide both inhibits the activity of the
CC angiotensin-converting enzyme and enhances the action of
CC bradykinin by inhibiting the kinases that inactivate it.
CC It acts as an indirect hypotensive agent.
DR PIR; JC0002; XAIVBH.
KW Hypotensive agent; Pyrrolidone carboxylic acid.

```

```

FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
SQ SEQUENCE 11 AA; 1112 MW; 30BABF1277686777 CRC64;

Query Match 79.2%; Score 38; DB 1; Length 11;
Best Local Similarity 75.0%; Pred. No. 0.92;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRPQGGP 8
DB 1 QGRPPGPP 8

RESULT 14
ID PMP3 MOUSE STANDARD; PRT; 296 AA.
AC P05143;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Proline-rich protein MP-3 (Fragment).
GN PRH1 OR PRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=86059475; PubMed=2999141;
RA Ann D.K., Carlson D.M.;
RT mouse structure and organization of a proline-rich protein gene of a
RL J. Biol. Chem. 260:15863-15872(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12100; AAA40005.1; -.
DR MGD; MGI:97773; Pphl.
DR NON_TER 1
SQ SEQUENCE 296 AA; 29521 MW; 7F146824B8AF3269 CRC64;

Query Match 79.2%; Score 38; DB 1; Length 296;
Best Local Similarity 85.7%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGGP 8
DB 282 GNPQGGP 288

RESULT 15
ID P12R RAT STANDARD; PRT; 416 AA.
AC P43253;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Prostaglandin receptor (Prostanoid IP receptor) (PGI receptor)
DE (Prostaglandin I2 receptor).
GN PTGIR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN RP SEQUENCE FROM N.A.
RT TISSUE=Lung;

```

```

RX MEDLINE=95101758; PubMed=7803522;
RA Sasaki Y., Ueui T., Tanaka I., Nakagawa O., Sando T., Takahashi T.,
RA Namba T., Narumiya S., Nakao K.;
RT "Cloning and expression of a cDNA for rat prostacyclin receptor.";
RL Biochim. Biophys. Acta 1224:601-605(1994).
CC 1- FUNCTION: RECEPTOR FOR PROSTACYCLIN (PROSTAGLANDIN I2 OR PGI2).
CC THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G-S PROTEINS WHICH
CC ACTIVATE ADENYLATE CYCLASE.
CC 1- SUBCELLULAR LOCATION: Integral membrane protein.
CC 1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D28966; BAA06091.1; -.
DR PIR; S52078; S52078.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR InterPro; IPR008365; ProstanoidR.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PRINTS; PR01788; PROSTANOIDR.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein; Lipoprotein;
KW Palmitate.
FT DOMAIN 1 45 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 46 67 1 (POTENTIAL).
FT DOMAIN 68 80 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 81 105 2 (POTENTIAL).
FT DOMAIN 106 123 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 124 144 3 (POTENTIAL).
FT DOMAIN 145 163 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 164 187 4 (POTENTIAL).
FT DOMAIN 188 215 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 216 237 5 (POTENTIAL).
FT DOMAIN 238 264 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 265 289 6 (POTENTIAL).
FT DOMAIN 290 302 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 303 323 7 (POTENTIAL).
FT DOMAIN 324 416 CYTOPLASMIC (POTENTIAL).
FT LIPID 336 336 S-palmitoyl cysteine [By similarity].
FT LIPID 339 339 S-palmitoyl cysteine [By similarity].
FT CARBOHYD 36 36 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 416 AA; 44662 MW; 03171B5ED21C4171 CRC64;

Query Match 79.2%; Score 38; DB 1; Length 416;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGGP 8
DB 6 GRPDGPP 12

Search completed: April 6, 2004, 16:07:58
Job time : 5.15888 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 28.9346 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-4  
Perfect score: 48  
Sequence: 1 RGRPQGP 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

SPTREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_thc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protist:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	256	16 Q7UYG7	Q7UYG7 rhodospirell
2	43	89.6	803	16 Q812W2	Q812W2 streptomyce
3	42	87.5	339	13 Q7ZUX2	Q7ZUX2 brachydanio
4	41	85.4	220	13 Q7N339	Q7N339 brachydanio
5	40	83.3	93	4 Q15215	Q15215 homo sapien
6	40	83.3	147	11 Q62104	Q62104 mus musculu
7	40	83.3	164	11 Q04105	Q04105 rattus norv
8	40	83.3	170	11 Q07610	Q07610 rattus norv
9	40	83.3	173	4 Q15214	Q15214 homo sapien
10	40	83.3	188	6 Q29427	Q29427 macaca fasc
11	40	83.3	202	11 Q04117	Q04117 rattus norv
12	40	83.3	238	4 Q00600	Q00600 homo sapien
13	40	83.3	297	4 Q16038	Q16038 homo sapien
14	40	83.3	309	4 Q04118	Q04118 homo sapien
15	40	83.3	338	4 Q86VA1	Q86VA1 homo sapien
16	40	83.3	358	4 Q08805	Q08805 homo sapien

17	40	83.3	382	4 Q00599	Q00599 homo sapien
18	40	83.3	395	11 Q9JIB1	Q9JIB1 rattus norv
19	40	83.3	683	16 Q7UR32	Q7UR32 rhodospirell
20	39	81.2	46	4 Q15218	Q15218 homo sapien
21	39	81.2	99	15 Q7ZB17	Q7ZB17 chimpanzee
22	39	81.2	209	4 Q8N811	Q8N811 homo sapien
23	39	81.2	697	4 Q60339	Q60339 homo sapien
24	39	81.2	804	4 Q8TEF5	Q8TEF5 homo sapien
25	39	81.2	1736	4 Q5I111	Q5I111 homo sapien
26	39	81.2	1857	4 Q9S153	Q9S153 homo sapien
27	38	79.2	66	12 Q911L8	Q911L8 white spot
28	38	79.2	69	12 Q91LD9	Q91LD9 white spot
29	38	79.2	144	13 Q7TLM3	Q7TLM3 bothrops ja
30	38	79.2	173	4 Q8HSU1	Q8HSU1 homo sapien
31	38	79.2	188	11 Q62106	Q62106 mus musculu
32	38	79.2	262	4 Q7Z7K5	Q7Z7K5 homo sapien
33	38	79.2	263	13 Q9PT52	Q9PT52 agkistrodon
34	38	79.2	295	11 Q07611	Q07611 rattus norv
35	38	79.2	317	11 Q62103	Q62103 mus musculu
36	38	79.2	333	4 Q96SD4	Q96SD4 homo sapien
37	38	79.2	367	4 Q8N6Q2	Q8N6Q2 homo sapien
38	38	79.2	367	4 Q8IXT2	Q8IXT2 homo sapien
39	38	79.2	370	11 Q8CGW9	Q8CGW9 mus musculu
40	38	79.2	504	11 Q91X93	Q91X93 mus musculu
41	38	79.2	564	4 Q9NXG8	Q9NXG8 homo sapien
42	38	79.2	811	16 Q9X8T0	Q9X8T0 streptomyce
43	38	79.2	842	4 Q8WXP9	Q8WXP9 homo sapien
44	38	79.2	1102	3 Q872D9	Q872D9 neurospora
45	37	77.1	18	4 Q9UCT9	Q9UCT9 homo sapien

#### ALIGNMENTS

#### RESULT 1

Q7UYG7 PRELIMINARY; PRT; 256 AA.

AC Q7UYG7; 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

GN RB624.

OS Rhodospirellula baltica.

OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;

OC Planctomycetaceae; Pirellula.

OX NCBI\_TaxID=117;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1;

RX MEDLINE=22735913; PubMed=12835416;

RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,

RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,

RA Schlesner H., Amann R., Reinhardt R.;

RT "Complete genome sequence of the marine planctomycete Pirellula sp. strain 1.1";

RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).

DR EMBL; BX294134; CAD71675.1; -.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 256 AA; 27003 MW; 2FAA06B5206EBB2F CRC64;

Query Match 100.0%; Score 48; DB 16; Length 256;

Best Local Similarity 100.0%; Pred. No. 1.2; 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0;

Qy 1 RGRPQGP 8

Db 160 RGRPQGP 167

#### RESULT 2

Q812W2 PRELIMINARY; PRT; 803 AA.

ID Q812W2

```

AC Q81ZW2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative serine/threonine protein kinase.
GN PKN27 OR SAV5424.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MA-4580 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Oseno T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RA "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MA-4580 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RA "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005042; BAC73136.1; -
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004674; P:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; P:protein-tyrosine kinase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR001128; Cytochrome P450.
DR InterPro; IPR007119; Prot kinase.
DR InterPro; IPR002290; Ser Thr pkinase.
DR InterPro; IPR008271; Ser Thr pkin AS.
DR InterPro; IPR001245; Tyr pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PDOM0001; Prot kinase; 1.
DR SMART; SM00220; S TKC; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00086; CYTOCHROME P450; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
KW Kinase; Serine/threonine-protein kinase; Complete proteome.
SQ SEQUENCE 803 AA; 82197 MW; C38C2BE0737FDB66 CRC64;

Query Match 89.6%; Score 43; DB 16; Length 803;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8
DB 195 GRPQGGP 201

RESULT 3
Q7ZUX2 PRELIMINARY; PRT; 339 AA.
AC Q7ZUX2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.

Query Match 85.4%; Score 41; DB 13; Length 220;
Best Local Similarity 75.0%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Body;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC046091; AAH46091.1; -
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
KW Hypothetical protein.
SQ SEQUENCE 339 AA; 35046 MW; 5BF0B1D6EB659148 CRC64;

Query Match 87.5%; Score 42; DB 13; Length 339;
Best Local Similarity 87.5%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPQGGP 8
DB 154 RGAQGGP 161

RESULT 4
Q7T339 PRELIMINARY; PRT; 220 AA.
AC Q7T339;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., McKernan K.J., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053265; AAH53265.1; -
KW Hypothetical protein.
SQ SEQUENCE 220 AA; 24769 MW; 2EC9DD7AB668C4AD CRC64;

Query Match 85.4%; Score 41; DB 13; Length 220;
Best Local Similarity 75.0%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```



QY 1 GRPQGGP 8  
 DB 7 RGKPGGP 14

## RESULT 5

Q15215 PRELIMINARY; PRT; 93 AA.  
 ID Q15215  
 AC Q15215;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Salivary proline-rich protein 1 (Fragment).  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84298476; PubMed=6089212;  
 RA Azen E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S.,  
 RA Maeda N., Vanin E.F., Carlson D.M., Smithies O.;  
 RT "Clones from the human gene complex coding for salivary proline-rich  
 RT proteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565(1984).  
 DR EMBL; K02576; AAA36503.1; -;  
 FT NON\_TER 1  
 SQ SEQUENCE 93 AA; 9170 MW; 04ADSF0797B31867 CRC64;

Query Match 83.3%; Score 40; DB 4; Length 93;  
 Best Local Similarity 85.7%; Pred. No. 10;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
 DB 20 GRPQGGP 26

## RESULT 6

Q62104 PRELIMINARY; PRT; 147 AA.  
 ID Q62104  
 AC Q62104;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE 15-kDa proline-rich salivary protein.  
 GN PEP15  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86033799; PubMed=3840480;  
 RA Clements S., Mehansho H., Carlson D.M.;  
 RT "Novel multigene families encoding highly repetitive peptide  
 RT sequences: Sequence analyses of rat and mouse proline-rich protein  
 RT cDNAs.";  
 RL J. Biol. Chem. 260:13471-13477(1985).  
 DR EMBL; M11900; AAA40000.1; -;  
 DR PIR; A29149; A29149.  
 DR MGD; MGI:1927481; Fp15.  
 SQ SEQUENCE 147 AA; 15107 MW; 96B3C4C7BC5FAC15 CRC64;

Query Match 83.3%; Score 40; DB 11; Length 147;  
 Best Local Similarity 85.7%; Pred. No. 17;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
 DB 66 GRPQGGP 72

## RESULT 7

Q04105 PRELIMINARY; PRT; 164 AA.  
 AC Q04105;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Salivary proline-rich protein.  
 GN RP13.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=91257817; PubMed=2045095;  
 RA Lin H.H., Ann D.K.;  
 RT "Molecular characterization of rat multigene family encoding proline-  
 RT rich proteins.";  
 RL Genomics 10:102-113(1991).  
 DR EMBL; M64792; AAA42063.1; -;  
 DR PIR; A30496; A30496.  
 SQ SEQUENCE 164 AA; 16806 MW; BD17EDFA699FFC03 CRC64;

Query Match 83.3%; Score 40; DB 11; Length 164;  
 Best Local Similarity 85.7%; Pred. No. 18;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
 DB 64 GRPQGGP 70

## RESULT 8

Q07610 PRELIMINARY; PRT; 170 AA.  
 ID Q07610  
 AC Q07610;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Proline-rich proteoglycan.  
 GN PRPGL  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Sprague-Dawley; TISSUE=Parotid gland;  
 RX MEDLINE=93388626; PubMed=8376404;  
 RA Castle A.M., Castle J.D.;  
 RT "Novel secretory proline-rich proteoglycans from rat parotid. Cloning  
 RT and characterization by expression in AtT-20 cells.";  
 RL J. Biol. Chem. 268:20490-20496(1993).  
 DR EMBL; L17317; AAA03073.1; -;  
 DR PIR; A48013; A48013.  
 SQ SEQUENCE 170 AA; 17840 MW; 85AF34C40B45AA03 CRC64;

Query Match 83.3%; Score 40; DB 11; Length 170;  
 Best Local Similarity 85.7%; Pred. No. 19;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
 DB 111 GRPQGGP 117

## RESULT 9

Q15214 PRELIMINARY; PRT; 173 AA.  
 ID Q15214  
 AC Q15214;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DE Salivary proline-rich protein 1 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84298176; PubMed=6089212;  
 RA Azen E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S.,  
 RA Maeda N., Vanin E.F., Carlson D.M., Smithies O.,  
 RT "Clones from the human gene complex coding for salivary proline-rich  
 RT proteins";  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565 (1984).  
 DR EMBL; K02575; AAA36502.1; -;  
 DR GO; GO:0005576; C:extracellular; NAS.  
 FT NON TER 1  
 FT NON TER 173 173  
 SQ SEQUENCE 173 AA; 17206 MW; BB4830335B77A7C0 CRC64;  
 Query Match 83.3%; Score 40; DB 4; Length 173;  
 Best Local Similarity 85.7%; Pred. No. 19;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GRPOGPP 8  
 Db 38 GKPOGPP 44  
 RESULT 10  
 ID Q29427 PRELIMINARY; PRT; 188 AA.  
 AC Q29427;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DE Proline-rich protein.  
 OS MNP4 OR MNP4P9.  
 GN Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92009216; PubMed=1916292;  
 RA Lin H.H., Kousvelari E.E., Ann D.K.;  
 RT "Sequence and expression of the Mnp4 gene encoding basic proline-rich  
 RT protein in macaque salivary glands";  
 RL Gene 104:219-226 (1991).  
 DR EMBL; M81322; AAA36905.1; -;  
 DR EMBL; M81321; AAA36904.1; -;  
 DR PIR; JH0481; JH0481.  
 SQ SEQUENCE 188 AA; 19135 MW; C8E6D0B7F4DEE504 CRC64;  
 Query Match 83.3%; Score 40; DB 6; Length 188;  
 Best Local Similarity 85.7%; Pred. No. 21;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GRPOGPP 8  
 Db 75 GKPOGPP 81  
 RESULT 11  
 ID Q04117 PRELIMINARY; PRT; 202 AA.  
 AC Q04117;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DE Salivary proline-rich protein 1 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91257817; PubMed=2045095;  
 RA Lin H.H., Ann D.K.;  
 RT "Molecular characterization of rat multigene family encoding proline-  
 RT rich proteins";  
 RL Genomics 10:102-113 (1991).  
 DR EMBL; M64791; AAA42066.1; -;  
 DR PIR; A39066; A39066.  
 SQ SEQUENCE 202 AA; 20801 MW; 9497F5B62F0065B9 CRC64;  
 Query Match 83.3%; Score 40; DB 11; Length 202;  
 Best Local Similarity 85.7%; Pred. No. 23;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GRPOGPP 8  
 Db 159 GKPOGPP 165

DE Salivary proline-rich protein.  
 GN R24.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RC MEDLINE=91257817; PubMed=2045095;  
 RA Lin H.H., Ann D.K.;  
 RT "Molecular characterization of rat multigene family encoding proline-  
 RT rich proteins";  
 RL Genomics 10:102-113 (1991).  
 DR EMBL; M64791; AAA42066.1; -;  
 DR PIR; A39066; A39066.  
 SQ SEQUENCE 202 AA; 20801 MW; 9497F5B62F0065B9 CRC64;  
 Query Match 83.3%; Score 40; DB 11; Length 202;  
 Best Local Similarity 85.7%; Pred. No. 23;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GRPOGPP 8  
 Db 159 GKPOGPP 165  
 RESULT 12  
 ID Q00600 PRELIMINARY; PRT; 238 AA.  
 AC Q00600;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Parotid 'o' protein (Fragment).  
 GN PRB4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96108975; PubMed=8554050;  
 RA Azen E.A., Amberger E., Fisher S., Prakobphol A., Niece R.L.;  
 RT "PRB1, PRB2, and PRB4 coded polymorphisms among human salivary  
 RT concanavalin-A binding, II-1, and Fc proline-rich proteins";  
 RL Am. J. Hum. Genet. 58:143-153 (1996).  
 DR EMBL; S80916; AAB50687.2; -;  
 FT NON TER 1  
 FT NON TER 1  
 SQ SEQUENCE 238 AA; 24005 MW; 4F44E947FF3A6C1 CRC64;  
 Query Match 83.3%; Score 40; DB 4; Length 238;  
 Best Local Similarity 85.7%; Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GRPOGPP 8  
 Db 25 GKPOGPP 31  
 RESULT 13  
 ID Q16038 PRELIMINARY; PRT; 297 AA.  
 AC Q16038;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE PRB1M protein (Fragment).  
 GN PRB1M.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;

```

KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 338 AA; 7F54B4E5AB002261 CRC64;
      1
Query Match 83.3%; Score 40; DB 4; Length 338;
Best Local Similarity 85.7%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 GRPQPPP 8
      :|:|:|
      :|:|:|
Db 62 GKPQPPP 68

Search completed: April 6, 2004, 16:14:39
Job time : 30.015 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 44.785 seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-4  
Perfect score: 48  
Sequence: 1 RGRPGPP 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	48	100.0	8	4	AAB48774	Aab48774 Human sal
2	48	100.0	9	4	AAB48775	Aab48775 Human sal
3	48	100.0	10	4	AAB48776	Aab48776 Human sal
4	48	100.0	17	4	AAB48783	Aab48783 Human sal
5	48	100.0	132	4	ABB38848	Abb38848 Peptide #
6	48	100.0	132	4	AAM32323	Aam32323 Peptide #
7	48	100.0	132	4	AM72058	Am72058 Human bon
8	48	100.0	132	4	AM59494	Ams59494 Human bra
9	48	100.0	132	4	ABG53742	Abg53742 Human liv
10	48	100.0	132	5	ABG41873	Abg41873 Human pep
11	48	100.0	149	6	ABR57423	AbR57423 Human NOV
12	48	100.0	154	6	ABR56769	AbR56769 Human sec
13	48	100.0	166	6	ADA83798	Ada83798 Human PRH
14	48	100.0	166	7	ADC98216	Adc98216 Human sal
15	43	89.6	44	2	AAW03557	Aaw03557 Human pro
16	43	89.6	45	1	AAE93320	Aap93320 P-C gene.
17	43	89.6	111	6	ABP75970	Abp75970 Human GEN
18	43	89.6	111	6	ABP76136	Abp76136 Human GEN
19	43	89.6	162	7	AAO30407	Aao30407 Human sec
20	43	89.6	267	7	AAO30408	Aao30408 Human sec
21	43	89.6	591	4	ABG15767	Abg15767 Novel hum
22	43	89.6	591	4	ABG18110	Abg18110 Novel hum
23	41	85.4	7	4	AAB48773	Aab48773 Human sal
24	41	85.4	106	6	ABP76137	Abp76137 Human GEN
25	41	85.4	106	6	ABP75971	Abp75971 Human GEN

26	40	83.3	18	2	AAW03556	Aaw03556 Octadecap
27	40	83.3	82	2	AAW50193	Aaw50193 Amino aci
28	40	83.3	117	6	AAO30249	Aao30249 Human Sap
29	40	83.3	124	2	AAW50192	Aaw50192 Amino aci
30	40	83.3	164	7	ADD47283	Add47283 Rat Prote
31	40	83.3	192	6	AAO30246	Aao30246 Human Sap
32	40	83.3	202	7	ADD48701	Add48701 Rat Prote
33	40	83.3	208	6	AAO30188	Aao30188 Human CP3
34	40	83.3	247	6	ABU11891	Abu11891 Human ABC
35	40	83.3	331	7	ADD45740	Add45740 Human Pro
36	40	83.3	335	4	AAE03634	Aae03634 Rat G-Pro
37	40	83.3	395	5	AAE14263	Aae14263 Rat NMUR2
38	40	83.3	1146	6	AAE34447	Aae34447 Human l1p
39	40	83.3	1165	4	ABG00133	Abg00133 Novel hum
40	39	81.2	41	6	AAO30248	Aao30248 Human Sap
41	39	81.2	49	3	AAO34155	Aao34155 Gene 13 h
42	39	81.2	112	4	AAO02189	Aao02189 Human pol
43	39	81.2	724	3	AAW70451	Aaw70451 Human mem
44	39	81.2	919	4	ABG07234	Abg07234 Novel hum
45	39	81.2	1857	3	AAW53970	Aaw53970 Human per

ALIGNMENTS

RESULT 1  
AAB48774  
ID AAB48774 standard; peptide; 8 AA.  
XX AAB48774;  
AC AAB48774;  
XX 09-MAR-2001 (first entry)  
DT Human saliva PRP-1 fragment (residues 106-113), SEQ ID NO:4.  
DE Human; PRP-1; proline-rich protein; saliva; dental caries;  
XX chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX Homo sapiens.  
OS WO200069890-A1.  
XX 23-NOV-2000.  
PD 11-MAY-2000; 2000WO-SE000930.  
PF 17-MAY-1999; 99SE-00001773.  
PR (STRO/) STROEMBERG N.  
XX (JOHA/) JOHANSSON I.  
PI Stroemberg N, Johansson I;  
XX WPI; 2001-031923/04.  
DR New oligopeptides comprising 2 arginine residues from degradation of  
XX proline-rich proteins, useful for preventing dental caries.  
PT Claim 4; Page 24; 36pp; English.  
CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
CC AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

CC derived oligopeptides of the invention

XX SQ Sequence 8 AA;

Query Match 100.0%; Score 48; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGGP 8  
| | | | |  
Db 1 RGRPQGGP 8

#### RESULT 2

AA48775  
ID AAB48775 standard; peptide; 9 AA.

XX AC

XX AAB48775;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 106-114), SEQ ID NO:5.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;

XX KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

XX KW oral bacterium; caries prevention.

XX OS Homo sapiens.

XX PN WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SB000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX PA (STRO/) STROEMBERG N.

XX PA (JOHA/) JOHANSSON I.

XX PI Stroemberg N, Johansson I;

XX DR WPI; 2001-031923/04.

XX PT New oligopeptides comprising 2 arginine residues from degradation of

XX PT proline-rich proteins, useful for preventing dental caries.

XX PS Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
AAB48783) which contain at least two arginine residues and which protect  
against dental caries. PRPs (proline-rich proteins) are salivary proteins  
encoded by six clustered genes on chromosome 12p13.2 and are potential  
determinants of a person's susceptibility to dental caries. PRPs are  
degraded by Actinomyces and Streptococcus species to small peptide  
fragments, these are metabolised by oral bacteria for nutritional  
purposes, with certain bacterial species generating ammonia via the  
catabolism of arginine. The peptides of the invention, being arginine-  
rich, can also be converted to ammonia by these bacteria. The ammonia  
thus formed raises the pH at the dental surface, thereby protecting the  
teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
derived oligopeptides of the invention

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 48; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGGP 8  
| | | | |  
Db 1 RGRPQGGP 8

#### RESULT 3

AA48776

ID AAB48776 standard; peptide; 10 AA.

XX AC AAB48776;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 106-115), SEQ ID NO:6.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;

XX KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

XX KW oral bacterium; caries prevention.

XX OS Homo sapiens.

XX PN WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SB000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX PA (STRO/) STROEMBERG N.

XX PA (JOHA/) JOHANSSON I.

XX PI Stroemberg N, Johansson I;

XX DR WPI; 2001-031923/04.

XX PT New oligopeptides comprising 2 arginine residues from degradation of

XX PT proline-rich proteins, useful for preventing dental caries.

XX PS Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
AAB48783) which contain at least two arginine residues and which protect  
against dental caries. PRPs (proline-rich proteins) are salivary proteins  
encoded by six clustered genes on chromosome 12p13.2 and are potential  
determinants of a person's susceptibility to dental caries. PRPs are  
degraded by Actinomyces and Streptococcus species to small peptide  
fragments, these are metabolised by oral bacteria for nutritional  
purposes, with certain bacterial species generating ammonia via the  
catabolism of arginine. The peptides of the invention, being arginine-  
rich, can also be converted to ammonia by these bacteria. The ammonia  
thus formed raises the pH at the dental surface, thereby protecting the  
teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
derived oligopeptides of the invention

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 48; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.42;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGGP 8  
| | | | |  
Db 1 RGRPQGGP 8

#### RESULT 4

AA48783

ID AAB48783 standard; peptide; 17 AA.

XX AC AAB48783;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 OS Homo sapiens.

XX WO200069890-A1.  
 XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.  
 XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.  
 XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;  
 XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX Claim 2; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention

XX Sequence 17 AA;

Query Match 100.0%; Score 48; DB 4; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 0.68; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 0;

Qy 1 RGRPQGGP 8  
 Db 8 RGRPQGGP 15

RESULT 5  
 ABB39848  
 ID ABB38848 standard; peptide; 132 AA.

XX AC ABB38848;

XX 04-FEB-2002 (first entry)

DE Peptide #6354 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human foetal liver.

XX Claim 27; SEQ ID NO 31483; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring  
 CC human gene expression in a sample derived from human foetal liver. The  
 CC single exon nucleic acid probes may be used for predicting, measuring and  
 CC displaying gene expression in samples derived from human foetal liver. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid probe  
 CC of the invention. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 132 AA;

Query Match 100.0%; Score 48; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 4.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQGGP 8

Db 88 RGRPQGGP 95

RESULT 6

AM32323  
 ID AAM32323 standard; protein; 132 AA.

XX AC AAM32323;

XX 17-OCT-2001 (first entry)

DE Peptide #6360 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human placenta.

XX Claim 27; SEQ ID NO 32592; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP: see AA131315-AA157546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 48; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGPP 8  
DB 88 RGRPOGPP 95

RESULT 7  
ID AAM72058 standard; protein; 132 AA.

XX AAM72058;  
XX AC  
XX DT 06-NOV-2001 (first entry)  
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32364.  
XX DE Human; bone marrow expressed exon; gene expression analysis; probe;  
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX OS Homo sapiens.  
XX PN WO200157276-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US000668.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PD WPI; 2001-488900/53.  
XX PF 30-JAN-2001; 2001WO-US000668.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PD WPI; 2001-488900/53.  
XX PF Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human bone marrow.

XX Example 4; SEQ ID NO 32364; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human bone marrow. They can be used to measure gene expression in bone marrow samples, which may enable the improved diagnosis and treatment of cancers such as lymphoma, leukaemia and myeloma. The present sequence is a protein encoded by one of the probes of the invention

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 48; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGPP 8  
DB 88 RGRPOGPP 95

RESULT 8  
ID AAM72058 standard; protein; 132 AA.

XX AAM72058;  
XX AC  
XX DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31599.  
XX DE Human; brain expressed exon; gene expression analysis; probe; microarray;  
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX OS Homo sapiens.  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US000667.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PD WPI; 2001-483446/52.  
XX PF Single exon nucleic acid probes for analyzing gene expression in human brains.

XX Example 4; SEQ ID NO 31599; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 48; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGPP 8  
DB 88 RGRPOGPP 95

RESULT 9  
ID ABG53742 standard; peptide; 132 AA.

XX ABG53742;  
XX AC  
XX DT 25-FEB-2003 (first entry)  
XX

DE Human liver peptide, SEQ ID NO 32390.  
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KW hypercholesterolaemia; coronary heart disease.  
XX  
OS Homo sapiens.  
XX  
XX WO200157273-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000664.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX  
XX 26-MAY-2000; 2000US-0207456P.  
XX  
XX 30-JUN-2000; 2000US-00608408.  
XX  
XX 03-AUG-2000; 2000US-00632366.  
XX  
XX 21-SEP-2000; 2000US-0234687P.  
XX  
XX 27-SEP-2000; 2000US-0236359P.  
XX  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488898/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human adult liver.  
XX  
XX Claim 27; SEQ ID NO 32390; 658pp; English.  
XX  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
XX measuring human gene expression in a sample derived from human adult  
XX liver, comprising one of 13109 defined nucleotide sequences given in the  
XX specification for complements/ fragments). The probe hybridises at high  
XX stringency to a nucleic acid molecule expressed in the human adult liver.  
XX (I) may be used for predicting, measuring and displaying gene expression  
XX in samples derived from human adult liver. The genes identified may be  
XX involved in genetic liver diseases such as cirrhosis,  
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
XX associated with coronary heart disease. ABG47348-ABG59930 represent human  
XX liver single exon encoded peptides of the invention. Note: The sequence  
XX information for this patent does not appear in the printed specification  
XX but was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 132 AA;  
XX  
XX Query Match 100.0%; Score 48; DB 4; Length 132;  
XX Best Local Similarity 100.0%; Pred. No. 4.4;  
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX Qy 1 RGRFQPPP 8  
XX |||||  
XX 88 RGRFQPPP 95  
XX  
XX RESULT 10  
XX ABG41873  
XX ID ABG41873 standard; peptide; 132 AA.  
XX  
XX AC ABG41873;  
XX  
XX 19-AUG-2002 (first entry)  
XX  
XX Human peptide encoded by genome-derived single exon probe SEQ ID 31538.  
XX  
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;  
XX chronic obstructive pulmonary disease; interstitial lung disease;  
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;  
XX tuberculous sclerosis; Gaucher's disease; Niemann-Pick disease;  
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
XX

KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW Primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.  
XX  
XX Homo sapiens.  
XX  
XX WO200186003-A2.  
XX  
XX 15-NOV-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000665.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX  
XX 26-MAY-2000; 2000US-0207456P.  
XX  
XX 30-JUN-2000; 2000US-00608408.  
XX  
XX 03-AUG-2000; 2000US-00632366.  
XX  
XX 21-SEP-2000; 2000US-0234687P.  
XX  
XX 27-SEP-2000; 2000US-0236359P.  
XX  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples.  
XX  
XX Claim 27; SEQ ID NO 31538; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human lung comprising single exon nucleic acid probes having one of  
XX 12614 nucleic acid sequences mentioned in the specification, or their  
XX complements or the 12387 open reading frames derived from the 12614  
XX probes. Also included are a microarray comprising the novel set of probes  
XX; the novel set of probes which hybridise at high stringency to a nucleic  
XX acid expressed in the human lung; measuring gene expression in a sample  
XX derived from human lung, comprising (a) contacting the array with a  
XX collection of detectably labeled nucleic acids derived from human lung  
XX mRNA, and (b) measuring the label detectably bound to each probe of the  
XX array; identifying exons in a eukaryotic genome, comprising (a)  
XX algorithmically predicting at least one exon from genomic sequences of  
XX the eukaryote; and (b) detecting specific hybridisation of detectably  
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
XX having a fragment identical to the predicted exon, the probe is included  
XX in the above mentioned microarray; assigning exons to a single gene,  
XX comprising (a) identifying exons from genomic sequence by the method  
XX above and (b) measuring the expression of each of the exons in several  
XX tissues and/or cell types using hybridisation to a single exon  
XX microarrays having a probe with the exon, where a common pattern of  
XX expression of the exons in the tissues and/or cell types indicates that  
XX the exons should be assigned to a single gene; a peptide comprising one  
XX of 12011 sequences, mentioned in the specification, or encoded by the  
XX probes/open reading frames (ORF). The probes are used for gene expression  
XX analysis, and for identifying exons in a gene, particularly using human  
XX lung derived mRNA and for the study of lung diseases such as asthma, lung  
XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
XX tuberculous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
XX Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
XX histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
XX Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
XX dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
XX present sequence is a peptide/protein encoded by a single exon probe of  
XX the invention. Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 132 AA;  
XX SQ



Query Match 100.0%; Score 48; DB 5; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 4.4; 0; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGPP 8  
 DB 88 RGRPOGPP 95  
 |||||

RESULT 11  
 ABR57423 standard; protein; 149 AA.

ID ABR57423  
 AC ABR57423;  
 DT 15-SEP-2003 (first entry)  
 DE Human NOV7 protein SEQ ID NO:24.

XX Human; NOVX; cytostatic; cardiant; antiinflammatory; immunosuppressive;  
 KW antiallergic; haemostatic; anti-HIV; antidiabetic; antiarteriosclerotic;  
 KW anorectic; antiaesthetic; nephrotropic; antiarthritic; hepatotropic;  
 KW neuroprotective; nontropic; antibacterial; virucide; antiparasitic;  
 KW relaxant; anticonvulsant; hypotensive; vasotropic; antiparkinsonian;  
 KW vulnerary; angiogenic; angiogenic; gene therapy; vaccine; cancer;  
 KW cardiomyopathy; atherosclerosis; hypertension; diabetes; inflammation;  
 KW autoimmune disorder; allergy; blood disorder; AIDS; obesity; asthma;  
 KW acquired immunodeficiency syndrome; nephropathy; cirrhosis; arthritis;  
 KW Alzheimer's disease; Parkinson's disease; Goitre; infection; stroke;  
 KW muscular dystrophy; epilepsy; wasting disorder; chromosome 12.

XX Homo sapiens.  
 OS  
 XX WO200294870-A2.  
 XX  
 XX 28-NOV-2002.  
 XX  
 XX 02-NOV-2001; 2001WO-US051580.  
 XX  
 XX 02-NOV-2000; 2000US-0245291P.  
 XX 02-NOV-2000; 2000US-0245317P.  
 XX 07-NOV-2000; 2000US-0246562P.  
 XX 08-NOV-2000; 2000US-0246871P.  
 XX 26-JAN-2001; 2001US-0264389P.  
 XX 26-JAN-2001; 2001US-0264423P.  
 XX 29-JAN-2001; 2001US-0264799P.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Grosse WM, Macdougall JR, Smithson G, Millet I, Stone DJ;  
 PI Gunther E, Ellerman K, Alsobrook JP, Lepley DM, Burgess CE;  
 PI Spytek KA, Edinger SR, Gangolli EA, Gorman L, Taupier RJ, Li L;  
 PI Guo X, Fernandes ER, Vernet CM, Tchernev VT, Casman SJ, Shenoy S;  
 PI Mishra V, Furtak K, Baumgartner JC, Colman SD;  
 XX WPI; 2003-140359/13.  
 DR N-PSDB; ACF03558.  
 XX  
 XX New NOVX polypeptide useful for preventing or treating NOVX-associated  
 PT disorders, e.g. cancer, cardiomyopathy, atherosclerosis or diabetes, and  
 PT in chromosome mapping, tissue typing or pharmacogenomics.  
 XX  
 XX Claim 1; Page 69; 346pp; English.

XX ACF03547 to ACF03570 encode the human NOVX proteins (I) given in ABR57412  
 CC to ABR57435. (I) have cytostatic, cardiant, antiinflammatory, nontropic,  
 CC immunosuppressive, antiallergic, haemostatic, antidiabetic,  
 CC antiarteriosclerotic, anorectic, antiaesthetic, nephrotropic, virucide,  
 CC antiarthritic, hepatotropic, neuroprotective, antibacterial, relaxant,  
 CC antiparasitic, anticonvulsant, hypotensive, vasotropic, antiparkinsonian,  
 CC vulnerary, angiogenic and angiogenic activities, and can be used in  
 CC gene therapy and vaccines. The NOVX polypeptides and their antibodies can  
 CC be used to determine the presence or absence of (I) in a sample. The NOVX

CC polypeptides, polynucleotides encoding them, and antibodies against them,  
 CC are useful in manufacturing a medicament for treating or preventing a  
 CC syndrome associated with a NOVX-associated disorder such as hypertension,  
 CC cardiomyopathy, atherosclerosis, cancer, diabetes, asthma, inflammation,  
 CC autoimmune disorders, allergies, blood disorders, obesity, acquired  
 CC immunodeficiency syndrome (AIDS), immunoglobulin (Ig)A nephropathy,  
 CC cirrhosis, arthritis, Alzheimer's disease, Parkinson's disease, goitre,  
 CC infections (e.g. bacterial, viral, parasitic), stroke, muscular  
 CC dystrophy, epilepsy, and other wasting disorders associated with chronic  
 CC diseases. ACF03571 to ACF03644 represent PCR primers and probes for NOVX  
 CC sequence, which are used in an example from the present invention

XX Sequence 149 AA;  
 SQ

Query Match 100.0%; Score 48; DB 6; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 4.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPOGPP 8  
 DB 105 RGRPOGPP 112  
 |||||

RESULT 12  
 ABR56769  
 ID ABR56769 standard; protein; 154 AA.

XX ABR56769;  
 AC ABR56769;  
 XX 30-JUL-2003 (first entry)  
 DT  
 XX Human secreted protein SECP-44 SEQ ID NO:44.

XX Human; secreted protein; SECP; cytostatic; antiarteriosclerotic;  
 KW anticonvulsant; nontropic; neuroprotective; cerebroprotective; anti-HIV;  
 KW antiallergic; antiinflammatory; thyromimetic; gene therapy; cancer;  
 KW cell proliferative disorder; atherosclerosis; neurological disorder;  
 KW epilepsy; Huntington's disease; stroke; immune disorder; allergy; AIDS;  
 KW inflammatory disorder; developmental disorder; hypothyroidism;  
 KW Cushing's syndrome; infection.

XX Homo sapiens.  
 OS  
 XX WO2003016506-A2.  
 XX  
 XX 27-FEB-2003.  
 XX  
 XX 15-AUG-2002; 2002WO-US027143.  
 XX  
 XX 17-AUG-2001; 2001US-0313249P.  
 XX 24-AUG-2001; 2001US-0314752P.  
 XX 07-SEP-2001; 2001US-0317818P.  
 XX 07-SEP-2001; 2001US-0317824P.  
 XX 21-SEP-2001; 2001US-0324040P.  
 XX 24-SEP-2001; 2001US-0324586P.  
 XX 02-NOV-2001; 2001US-0343980P.  
 XX 28-NOV-2001; 2001US-0343229P.  
 XX 13-FEB-2002; 2002US-0357002P.  
 XX 06-MAR-2002; 2002US-0362439P.  
 XX 19-MAR-2002; 2002US-0366041P.  
 XX 30-APR-2002; 2002US-0376988P.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX Tang TY, Warren BA, Gietzen KJ, Lal PG, Yue H, Honchell CD;  
 PI Lehr-Mason PM, Burford N, Xu Y, Baughn MB, Duggan BM, Tran UK;  
 PI Lee EN, Forsythe IJ, Richardson TW, Lee S, Thangavelu K, Yue H;  
 PI Emerling BM, Walla NK, Azimzai Y, Sanjanwala B, Hafalia AJA;  
 PI Borowsky ML, Nguyen DB, Ison CH, Aströmoff A, Ding L, Lee SY;  
 PI Becha SD, Ramkumar J, Gandhi AR, Jin P, Fu GK, Swarnakar A;  
 XX WPI; 2003-278569/27.  
 DR N-PSDB; ACC79069.  
 DR

XX New human secreted proteins (SECP), useful for diagnosing, treating and  
PT preventing diseases or conditions associated with the aberrant SECP  
PT expression e.g. cancer, AIDS, atherosclerosis, Huntington's disease,  
PT stroke, infections.

XX Claim 1; Page 222; 286pp; English.

XX ACC79026 to ACC79105 encode the human secreted proteins (I) given in  
CC ABR56726 to ABR56805, designated SECP-1 to SECP-80. SECP sequences can  
CC have cytostatic, antiarteriosclerotic, anticonvulsant, antiinflammatory,  
CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic and  
CC thymimetic activities, and can be used in gene therapy. The SECP  
CC proteins and polynucleotides can be used in diagnosing, treating and  
CC preventing diseases or conditions associated with the decreased  
CC expression or overexpression of SECP, such as cell proliferative (e.g.  
CC cancer, atherosclerosis), neurological (e.g. AIDS, allergies) and  
CC developmental (e.g. hypothyroidism, Cushing's syndrome) disorders, or  
CC infections. They are also useful in assessing the effects of exogenous  
CC compounds on the expression of nucleic acid and amino acid sequences of  
CC SECP. The SECP or its fragments are useful in screening compounds for  
CC effectiveness as agonist or antagonist of the polypeptides, or in  
CC altering the expression of the target polynucleotide and compounds that  
CC specifically bind to or modulate the activity of the polypeptide

XX Sequence 154 AA;  
SQ Query Match 100.0%; Score 48; DB 6; Length 154;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQGGP 8  
Db 110 RGRPQGGP 117

RESULT 13  
ADA83798  
ID ADA83798 standard; protein; 166 AA.  
XX  
AC ADA83798;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Human PRH2 protein.  
XX  
KW human; marker; expressed sequence tag; EST; arabidopsis; tumour;  
KW stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;  
KW vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO2002103028-A2.  
XX  
PD 27-DEC-2002.  
XX  
PF 30-MAY-2002; 2002WO-IB004189.  
XX  
PR 30-MAY-2001; 2001US-0293999P.  
PR 22-OCT-2001; 2001US-0330457P.  
PR 19-FEB-2002; 2002US-0357144P.  
XX  
PA (BIOM-) BIOMEDICAL CENT.  
XX  
PI Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;  
XX  
DR WPI; 2003-175241/17.  
XX  
DR N-PSDB; ADA83797.  
XX  
PT Determining if a nucleic acid is a marker for a phenotype/cell type of  
PT interest, by global comparison of expressed sequence tags known to be  
PT expressed in the phenotype/cell type with all ESTs expressed in normal

PT tissue.  
XX Claim 29; Page 191-192; 516pp; English.  
XX  
CC The invention relates to a novel method for determining if a nucleic acid  
CC is a marker for a predetermined phenotype/cell type of interest from a  
CC biological species. The method comprises performing a global comparison  
CC of a group of expressed sequence tags (ESTs) known to be expressed in the  
CC phenotype/cell type of interest with all ESTs expressed in normal tissue  
CC in order to identify ESTs that are preferentially expressed in the  
CC phenotype/cell type of interest. A method of the invention is useful for  
CC determining whether a nucleic acid is a marker for a predetermined  
CC phenotype or cell type of interest from a biological species, preferably  
CC Arabidopsis or human. The cell type of interest is an abnormal cell such  
CC as a tumour cell, and the predetermined phenotype is a stress-induced  
CC phenotype such as hyperosmotic stress or high salt conditions. A method  
CC of the invention is also useful for determining the progression of colon  
CC cancer in a human, for detecting a tumour cell, and for regulating or  
CC preventing the growth of a tumour cell. An antibody of the invention is  
CC useful for detecting the absence or presence of peptides encoded by  
CC tumour-associated markers. A polypeptide of the invention is useful as an  
CC immunogen for vaccinating an animal. The present sequence represents a  
CC tumour-associated antigen of the invention.

XX Sequence 166 AA;  
SQ Query Match 100.0%; Score 48; DB 6; Length 166;  
Best Local Similarity 100.0%; Pred. No. 5.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQGGP 8  
Db 122 RGRPQGGP 129

RESULT 14  
ADC98216  
ID ADC98216 standard; protein; 166 AA.  
XX  
AC ADC98216;  
XX  
DT 01-JAN-2004 (first entry)  
XX  
DE Human salivary acidic proline-rich phosphoprotein (PRP).  
XX  
KW Human; salivary acidic proline-rich phosphoprotein; PRP; autoantigen;  
KW atopic dermatitis-inducing protein; salivary gland; IgE autoantibody;  
KW immunoglobulin E; mast cell activation; basophil activation; diagnosis;  
KW risk assessment; sensitisation remedy; dermatological; antiallergic;  
KW antiinflammatory.  
XX  
OS Homo sapiens.  
XX  
PN WO2003084991-A1.  
XX  
PD 16-OCT-2003.  
XX  
PF 04-APR-2003; 2003WO-JP004325.  
XX  
PR 08-APR-2002; 2002JP-00105425.  
XX  
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX  
PI Hide M, Yamamoto S, Tanaka T, Koro O;  
XX  
DR WPI; 2003-833567/77.  
XX  
DR N-PSDB; ADC98215.  
XX  
PT Atopic dermatitis-inducing proteins, applicable in diagnosis of including  
PT risk of onset, and in developing sensitization remedies for the disease.  
XX Claim 4; SEQ ID NO 2; 43pp; Japanese.  
XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 31.4019 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-4

Perfect score: 48  
Sequence: 1 RGRPQGP 8

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA.\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	132	9	US-09-864-761-43644
2	48	100.0	166	14	US-10-157-031-80
3	43	89.6	803	14	US-10-156-761-12958
4	41	85.4	149	12	US-10-424-599-171613
5	40	83.3	159	12	US-10-424-599-212996
6	39	81.2	92	14	US-10-029-386-29755
7	39	81.2	209	15	US-10-108-260A-4370
8	38	79.2	150	12	US-10-424-599-146985
9	38	79.2	842	12	US-10-112-944-373
10	38	79.2	5065	14	US-10-171-311-52
11	37	77.1	24	9	US-09-802-674-2
12	37	77.1	28	14	US-10-262-017-5
13	37	77.1	45	9	US-09-975-143-30
14	37	77.1	52	12	US-10-424-599-277859
15	37	77.1	53	12	US-10-424-599-171786

16	37	77.1	76	12	US-10-424-599-204459	Sequence 204459,
17	37	77.1	87	12	US-10-276-774-1721	Sequence 1721, Ap
18	37	77.1	114	10	US-09-764-891-5325	Sequence 5325, Ap
19	37	77.1	114	10	US-09-942-024-19	Sequence 19, Appl
20	37	77.1	114	10	US-09-942-098-19	Sequence 19, Appl
21	37	77.1	114	10	US-10-205-428-480	Sequence 480, App
22	37	77.1	121	12	US-10-424-599-271545	Sequence 271545,
23	37	77.1	127	12	US-10-424-599-223622	Sequence 223622,
24	37	77.1	133	12	US-10-424-599-216185	Sequence 216185,
25	37	77.1	210	14	US-10-029-386-33437	Sequence 33437, A
26	37	77.1	268	14	US-10-106-698-6399	Sequence 6399, Ap
27	37	77.1	274	9	US-09-850-887-4	Sequence 850, Appl
28	37	77.1	281	14	US-10-262-017-5	Sequence 5, Appli
29	37	77.1	323	9	US-09-802-674-2	Sequence 2, Appli
30	37	77.1	323	9	US-09-833-263-1064	Sequence 1064, Ap
31	37	77.1	323	9	US-09-833-263-1064	Sequence 1064, Ap
32	37	77.1	323	9	US-09-833-263-1064	Sequence 2, Appli
33	37	77.1	323	9	US-09-833-263-1064	Sequence 2, Appli
34	37	77.1	323	9	US-09-833-263-1064	Sequence 110, App
35	37	77.1	323	13	US-10-025-380-1064	Sequence 1064, Ap
36	37	77.1	323	14	US-10-235-994-30	Sequence 2, Appli
37	37	77.1	323	14	US-10-235-994-30	Sequence 30, Appl
38	37	77.1	323	14	US-10-376-133-17	Sequence 17, Appl
39	37	77.1	329	9	US-09-802-674-13	Sequence 13, Appl
40	37	77.1	351	14	US-10-106-698-4502	Sequence 4502, Ap
41	37	77.1	440	12	US-10-425-114-39512	Sequence 39512, A
42	37	77.1	476	14	US-10-156-761-8693	Sequence 8693, Ap
43	37	77.1	551	11	US-09-764-875-892	Sequence 892, App
44	37	77.1	641	10	US-09-959-987-2	Sequence 2, Appli
45	37	77.1	647	9	US-09-925-299-1002	Sequence 1002, Ap
			647	10	US-09-925-299-1002	Sequence 1002, Ap

## ALIGNMENTS

### RESULT 1

US-09-864-761-43644  
; Sequence 43644, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aescmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43644
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC00518.17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
; OTHER INFORMATION: EST HUMAN HIT: BF088785.1, EVALUE 1.00e-01
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
US-09-864-761-43644

Query Match      100.0%; Score 48; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPGQPP 8
DB 88 RGRPGQPP 95

RESULT 2
US-10-157-031-80
; Sequence 80, Application US/10157031
; Publication No. US2003010890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Rankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Kravskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-80

Query Match      100.0%; Score 48; DB 14; Length 166;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPGQPP 8
DB 122 RGRPGQPP 129

RESULT 3
US-10-156-761-12958
; Sequence 12958, Application US/10156761
```

```
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-282
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12958
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12958

Query Match      89.6%; Score 43; DB 14; Length 803;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPGQPP 8
DB 195 GRPGQPP 201

RESULT 4
US-10-424-599-171613
; Sequence 171613, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 171613
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_125982C.1.pap
US-10-424-599-171613

Query Match      85.4%; Score 41; DB 12; Length 149;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGQPP 8
DB 55 RGRPGQPP 62

RESULT 5
US-10-424-599-212996
; Sequence 212996, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
```

; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 212996  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(159)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_34360C.1.pep  
US-10-424-599-212996

Query Match 83.3%; Score 40; DB 12; Length 159;  
Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
|||:||||  
DB 147 GRPEGPP 153

RESULT 6  
US-10-029-386-29755  
; Sequence 29755, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 29755  
; LENGTH: 92  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO CH17.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.84  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.71  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.76  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.61  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.99  
; OTHER INFORMATION: SWISSPROT HIT: O35569, EVALUATE 1.10e-02

US-10-029-386-29755  
Query Match 81.2%; Score 39; DB 14; Length 92;  
Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGP 8  
|||:||||  
DB 47 GRPRGPP 53

RESULT 7  
US-10-108-260A-4370  
; Sequence 4370, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20040005560A1 full length cDNA

; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4370  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-4370

Query Match 81.2%; Score 39; DB 15; Length 209;  
Best Local Similarity 87.5%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGGP 8  
|||:||||  
DB 119 RGRPGHP 126

RESULT 8  
US-10-424-599-146985  
; Sequence 146985, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 146985  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_103748C.1.pep  
US-10-424-599-146985

Query Match 79.2%; Score 38; DB 12; Length 150;  
Best Local Similarity 75.0%; Pred. No. 3.1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGGP 8  
|||:||||  
DB 104 RGRPGGP 111

RESULT 9  
US-10-112-944-373  
; Sequence 373, Application US/10112944  
; Publication No. US20040048249A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Yang, Yongzhong  
; APPLICANT: Weng, Gezhi  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Ghosh, Malabika  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wang, Zhiwei  
; TITLE OF INVENTION: No. US20040048249A1el Nucleic Acids and  
; FILE REFERENCE: 80SA  
; CURRENT APPLICATION NUMBER: US/10/112,944

DB 450 GEPQPP 456

RESULT 11

US-09-802-674-7

; Sequence 7, Application US/09802674

; Patent No. US20020042088A1

; GENERAL INFORMATION:

; APPLICANT: Macina, Roberto A

; APPLICANT: Piderit, Alejandra

; APPLICANT: Sun, Yongming

; TITLE OF INVENTION: Method of Diagnosing, Monitoring, Staging, and

; TITLE OF INVENTION: Treating Gastrointestinal Cancer

; FILE REFERENCE: DEX-0142

; CURRENT APPLICATION NUMBER: US/09/802,674

; CURRENT FILING DATE: 2001-03-09

; PRIOR APPLICATION NUMBER: 60/188,061

; PRIOR FILING DATE: 2000-03-09

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 7

; LENGTH: 24

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-802-674-7

Query Match 79.2%; Score 38; DB 12; Length 842;

Best Local Similarity 85.7%; Pred. No. 1.4e+03;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQPP 8

Db 456 GRPMGPP 462

RESULT 10

US-10-171-311-52

; Sequence 52, Application US/10171311

; Publication No. US20030087270A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Glatt, Karen

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Hoerish, Sebastian

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY

; TITLE OF INVENTION: OF CERVICAL CANCER

; FILE REFERENCE: MRI-035

; CURRENT APPLICATION NUMBER: US/10/171,311

; CURRENT FILING DATE: 2002-06-12

; PRIOR APPLICATION NUMBER: US 60/298,159

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/298,155

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/335,936

; PRIOR FILING DATE: 2001-11-14

; NUMBER OF SEQ ID NOS: 238

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 52

; LENGTH: 5065

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-171-311-52

Query Match 79.2%; Score 38; DB 14; Length 5065;

Best Local Similarity 85.7%; Pred. No. 7e+03;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQPP 8

Db 456 GRPMGPP 462

DB 450 GEPQPP 456

RESULT 12

US-10-262-017-3

; Sequence 3, Application US/10262017

; Publication No. US20030109464A1

; GENERAL INFORMATION:

; APPLICANT: Huflejt, Margaret E.

; APPLICANT: Mossine, Valeri V.

; APPLICANT: Croft, Michael

; TITLE OF INVENTION: GALECTINS-1 AND -4 IN TUMOR DEVELOPMENT

; FILE REFERENCE: 511762000100

; CURRENT APPLICATION NUMBER: US/10/262,017

; CURRENT FILING DATE: 2002-09-30

; PRIOR APPLICATION NUMBER: 60/326,137

; PRIOR FILING DATE: 2001-09-28

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 28

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-262-017-3

Query Match 77.1%; Score 37; DB 9; Length 24;

Best Local Similarity 100.0%; Pred. No. 83;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RPQPP 8

Db 3 RPQPP 8

RESULT 13

US-09-975-143-30

; Sequence 30, Application US/09975143

; Patent No. US2002015513A1

; GENERAL INFORMATION:

; APPLICANT: HSU, Daniel, K.

; APPLICANT: LIU, Fu-Tong

Query Match 77.1%; Score 37; DB 14; Length 28;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RPQPP 8

Db 6 RPQPP 11

; APPLICANT: DOWLING, Christopher, A.  
; TITLE OF INVENTION: GALECTIN EXPRESSION IS INDUCED IN  
; TITLE OF INVENTION: CIRRHOTIC LIVER AND HEPATOCELLULAR CARCINOMA  
; FILE REFERENCE: DANHSU.001C1  
; CURRENT APPLICATION NUMBER: US/09/975.143  
; CURRENT FILING DATE: 2003-10-10  
; PRIOR APPLICATION NUMBER: FC7/US00/08561  
; PRIOR FILING DATE: 2000-03-29  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 45  
; TYPE: PRT  
; ORGANISM: human  
US-09-975-143-30

Query Match 77.1%; Score 37; DB 9; Length 45;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RPOGPP 8  
|||  
Db 37 RPOGPP 42

RESULT 14  
US-10-424-599-277859  
; Sequence 277859, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(S3223)B  
; CURRENT APPLICATION NUMBER: US/10/424.599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 277859  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_9292C.1.pep  
US-10-424-599-277859

Query Match 77.1%; Score 37; DB 12; Length 52;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGQPP 8  
|||  
Db 1 RGRPGQPP 8

RESULT 15  
US-10-424-599-171786  
; Sequence 171786, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(S3223)B  
; CURRENT APPLICATION NUMBER: US/10/424.599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 171786

; LENGTH: 53  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_126137C.1.pep  
US-10-424-599-171786

Query Match 77.1%; Score 37; DB 12; Length 53;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGQPP 8  
|||  
Db 40 RGRPGQPP 47

Search completed: April 6, 2004, 17:05:52  
Job time : 31.4019 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 11.7383 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-4

Perfect score: 48

Sequence: 1 RGRPGPP 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/iaa/PCUTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	53.8	423	4	US-09-489-039A-7834
2	42	87.5	391	4	US-09-252-991A-20316
3	41	85.4	366	4	US-09-252-991A-17018
4	40	83.3	82	2	US-08-925-237-4
5	40	83.3	124	2	US-08-925-237-2
6	40	83.3	977	4	US-09-252-991A-16655
7	38	79.2	160	4	US-09-252-991A-27129
8	37	77.1	135	4	US-09-252-991A-21466
9	37	77.1	238	4	US-09-252-991A-23387
10	37	77.1	291	4	US-09-252-991A-19638
11	37	77.1	311	3	US-08-911-423-8
12	37	77.1	323	3	US-08-946-914-2
13	37	77.1	323	3	US-09-224-110-16
14	37	77.1	323	4	US-09-656-450-2
15	37	77.1	323	5	PCT-US95-07289-16
16	37	77.1	446	4	US-09-252-991A-27110
17	37	77.1	600	4	US-09-252-991A-28916
18	37	77.1	617	4	US-09-252-991A-29507
19	37	77.1	691	4	US-09-252-991A-16809
20	37	77.0	38	4	US-09-149-476-605
21	36	75.0	58	4	US-09-904-615-157
22	36	75.0	64	4	US-09-370-838-74
23	36	75.0	73	1	US-08-355-888A-11
24	36	75.0	73	1	US-08-693-897-11
25	36	75.0	73	3	US-08-693-896-11
26	36	75.0	73	4	US-09-357-914-11
27	36	75.0	73	4	US-09-357-914-11

28	36	75.0	114	4	US-09-252-991A-25597	Sequence 25597, A
29	36	75.0	132	4	US-09-252-991A-22177	Sequence 22177, A
30	36	75.0	168	1	US-08-451-947-6	Sequence 6, Appl
31	36	75.0	168	2	US-08-424-826A-6	Sequence 6, Appl
32	36	75.0	168	3	US-08-928-694-6	Sequence 6, Appl
33	36	75.0	168	4	US-08-450-842-6	Sequence 6, Appl
34	36	75.0	168	4	US-08-451-390-6	Sequence 6, Appl
35	36	75.0	168	5	PCT-US91-06950-6	Sequence 6, Appl
36	36	75.0	199	4	US-09-252-991A-27546	Sequence 27546, A
37	36	75.0	210	1	US-08-451-947-2	Sequence 2, Appl
38	36	75.0	210	2	US-08-424-826A-2	Sequence 2, Appl
39	36	75.0	210	3	US-08-928-694-2	Sequence 2, Appl
40	36	75.0	210	4	US-08-450-842-2	Sequence 2, Appl
41	36	75.0	210	4	US-08-451-390-2	Sequence 2, Appl
42	36	75.0	215	1	PCT-US91-06950-2	Sequence 2, Appl
43	36	75.0	215	5	US-07-796-106-23	Sequence 23, Appl
44	36	75.0	239	4	US-09-252-991A-26949	Sequence 26949, A
45	36	75.0	263	4	US-09-252-991A-16906	Sequence 16906, A

## ALIGNMENTS

### RESULT 1

US-09-489-039A-7834

; Sequence 7834, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 7834

; LENGTH: 423

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-7834

Query Match 93.8%; Score 45; DB 4; Length 423;

Best Local Similarity 87.5%; Pred. No. 15;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPGPP 8

DB 194 RGRPGPP 201

### RESULT 2

US-09-252-991A-20316

; Sequence 20316, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 20316

; LENGTH: 391

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20316

```
Query Match      87.5%; Score 42; DB 4; Length 391;
Best Local Similarity 87.5%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RGRPGQPP 8
    |||||
Db 105 RGRPGAGPP 112

RESULT 3
US-09-252-991A-17018
; Sequence 17018, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17018
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17018

Query Match      85.4%; Score 41; DB 4; Length 366;
Best Local Similarity 87.5%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RGRPGQPP 8
    |||||
Db 135 RGRPGGPP 142

RESULT 4
US-08-925-237-4
; Sequence 4, Application US/08925237
; Patent No. 5981720
; GENERAL INFORMATION:
; APPLICANT: Azen, Edwin A.
; TITLE OF INVENTION: Human Salivary Proteins And Fragments
; TITLE OF INVENTION: Thereof Having Alpha-Glucosidase Inhibitory Activity
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kent Barta
; STREET: 100 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: USA
; ZIP: 53202-4108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,237
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Barta, Kent S.
; REGISTRATION NUMBER: 29,042
; REFERENCE/DOCKET NUMBER: 96429/9002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 257-3501
; TELEFAX: (608) 283-2275
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-925-237-2

Query Match      83.3%; Score 40; DB 2; Length 124;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPGQPP 8
    |||||
Db 4 GRPGQPP 10

RESULT 6
US-09-252-991A-16655
; Sequence 16655, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Azen, Edwin A.
; TITLE OF INVENTION: Human Salivary Proteins And Fragments
; TITLE OF INVENTION: Thereof Having Alpha-Glucosidase Inhibitory Activity
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kent Barta
; STREET: 100 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: USA
; ZIP: 53202-4108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,237
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Barta, Kent S.
; REGISTRATION NUMBER: 29,042
; REFERENCE/DOCKET NUMBER: 96429/9002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 257-3501
```

```
TELEFAX: (608) 283-2275
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 82 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-925-237-4

Query Match      83.3%; Score 40; DB 2; Length 82;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPGQPP 8
    |||||
Db 4 GRPGQPP 10

RESULT 5
US-08-925-237-2
; Sequence 2, Application US/08925237
; Patent No. 5981720
; GENERAL INFORMATION:
; APPLICANT: Azen, Edwin A.
; TITLE OF INVENTION: Human Salivary Proteins And Fragments
; TITLE OF INVENTION: Thereof Having Alpha-Glucosidase Inhibitory Activity
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kent Barta
; STREET: 100 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: USA
; ZIP: 53202-4108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,237
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Barta, Kent S.
; REGISTRATION NUMBER: 29,042
; REFERENCE/DOCKET NUMBER: 96429/9002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 257-3501
; TELEFAX: (608) 283-2275
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-925-237-2

Query Match      83.3%; Score 40; DB 2; Length 124;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPGQPP 8
    |||||
Db 4 GRPGQPP 10

RESULT 6
US-09-252-991A-16655
; Sequence 16655, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
```

```
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16655
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16655

Query Match      83.3%; Score 40; DB 4; Length 977;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 RGRPGQPP 8
      |||||
Db      490 RGIQPGPP 497

RESULT 7
US-09-252-991A-27129
; Sequence 27129, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27129
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27129

Query Match      79.2%; Score 38; DB 4; Length 160;
Best Local Similarity 75.0%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 RGRPGQPP 8
      |||||
Db      17 RSRPRGPP 24

RESULT 8
US-09-252-991A-21466
; Sequence 21466, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
```

```
; SEQ ID NO 21466
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21466

Query Match      77.1%; Score 37; DB 4; Length 135;
Best Local Similarity 75.0%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 RGRPGQPP 8
      |||||
Db      12 RGRPSAPP 19

RESULT 9
US-09-252-991A-23387
; Sequence 23387, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23387
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23387

Query Match      77.1%; Score 37; DB 4; Length 238;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 RGRPGQPP 8
      |||||
Db      43 RERPEGPP 50

RESULT 10
US-09-252-991A-19638
; Sequence 19638, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19638
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19638

Query Match      77.1%; Score 37; DB 4; Length 291;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 RGRPGQPP 8
```

```

Db      232  RGRPEQP 239
||||:|
RESULT 11
US-08-911-423-8
; Sequence 8, Application US/08911423
; Patent No. 6111090
; GENERAL INFORMATION:
; APPLICANT: Gorman, Daniel M.
; APPLICANT: Randall, Troy D.
; APPLICANT: Zlotnick, Albert
; TITLE OF INVENTION: MAMMALIAN CELL SURFACE ANTIGENS; RELATED
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,423
; FILING DATE: 14-AUG-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,419
; FILING DATE: 16-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/027,901
; FILING DATE: 07-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0612K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-852-9196
; TELEFAX: 650-496-1200
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-911-423-8
Query Match      77.1%; Score 37; DB 3; Length 311;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2  GRPQGGP 8
|||||
Db      260  GRPQGGP 266

RESULT 12
US-08-469-667-16
; Sequence 16, Application US/08469667
; Patent No. 5733748
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,

```

```

; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,667
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fetraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 323 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-469-667-16
Query Match      77.1%; Score 37; DB 1; Length 323;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3  RPOGPP 8
|||||
Db      156  RPOGPP 161

RESULT 13
US-08-946-914-2
; Sequence 2, Application US/08946914
; Patent No. 6027916
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kassler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,914
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,093
; FILING DATE: 09-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 323 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-946-914-2

Query Match 77.1%; Score 37; DB 3; Length 323;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RPOGPP 8  
|||  
Db 156 RPOGPP 161

## RESULT 14

US-09-224-110-16  
; Sequence 16, Application US/09224110  
; Patent No. 6337195  
; GENERAL INFORMATION:  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Rosen, Craig  
; TITLE OF INVENTION: Colon Specific Genes and Proteins  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
; ADDRESSEE: Stewart & Olstein  
; STREET: 6 Becker Farm Road  
; CITY: Roseland  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07068-1739  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/224,110  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/469,667  
; FILING DATE: 06-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Feriayo, Gregory D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-435  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 323 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-224-110-16

Query Match 77.1%; Score 37; DB 4; Length 323;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RPOGPP 8  
|||  
Db 156 RPOGPP 161

## RESULT 15

US-09-656-450-2  
; Sequence 2, Application US/09656450  
; Patent No. 6488788  
; GENERAL INFORMATION:  
; APPLICANT: Ni, Jian  
; APPLICANT: Gentz, Reiner L.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides  
; FILE REFERENCE: 1488-0560003  
; CURRENT APPLICATION NUMBER: US/09/656,450  
; CURRENT FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: US 09/263,689  
; PRIOR FILING DATE: 1999-03-05  
; PRIOR APPLICATION NUMBER: US 08/946,914  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: US 60/028,093  
; PRIOR FILING DATE: 1996-10-09  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-656-450-2

Query Match 77.1%; Score 37; DB 4; Length 323;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RPOGPP 8  
|||  
Db 156 RPOGPP 161

Search completed: April 6, 2004, 16:19:39  
Job time : 11.7383 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 10.0935 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-5  
Perfect score: 53  
Sequence: 1 KGRPGPPQ 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	166	PIHUSC	salivary proline-r
2	53	100.0	166	B25372	salivary proline-r
3	53	100.0	171	A27307	proline-rich phosph
4	45	84.9	128	D33355	basic proline-rich
5	45	84.9	188	JH0481	basic proline-rich
6	45	84.9	251	PIHUPF	salivary proline-r
7	45	84.9	310	PIHUSD	salivary proline-r
8	45	84.9	392	PIHUB6	salivary proline-r
9	43	81.1	188	D29149	proline-rich prote
10	43	81.1	212	B34298	proline-rich prote
11	43	81.1	295	B48013	proline-rich prote
12	43	81.1	317	A28996	proline-rich prote
13	42	79.2	170	A48013	acidic proline-ric
14	42	79.2	206	PIRT3	hypothetical prote
15	42	79.2	797	S53590	synaptobrevin - lo
16	41	77.4	125	S40153	SREBP cleavage act
17	41	77.4	1276	T18526	basic proline-rich
18	40	75.5	76	C38355	proline-rich prote
19	40	75.5	117	D40750	proline-rich prote
20	40	75.5	147	A29149	proline-rich prote
21	40	75.5	164	A30496	proline-rich prote
22	40	75.5	172	B29149	proline-rich prote
23	40	75.5	204	A39066	proline-rich prote
24	40	75.5	223	A42817	proline-rich prote
25	40	75.5	278	B39066	proline-rich prote
26	40	75.5	309	S10889	cyclin - fission y
27	40	75.5	325	T40529	hypothetical prote
28	40	75.5	1736	T00391	embryonic abundant
29	39	73.6	266	T09281	

ALIGNMENTS

RESULT 1

PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human  
N:Alternate names: salivary acidic proline-rich protein PRH2  
N:Contains: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotein  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1981 #sequence revision 12-Apr-1996 #text change 08-Dec-2000  
C:Accession: A25372; A19803; B57868; A92277; A92254; A94425; A91954; S02564; S02563; JF  
R:Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
J. Biol. Chem. 260, 1123-1130, 1985  
A:Title: Differential RNA splicing and post-translational cleavages in the human saliv  
A:Reference number: A92492; MUID:85289325; PMID:2993301  
A:Accession: A25372

A:Molecule type: mRNA  
A:Residues: 1-166 <MBE>  
A:Cross-references: GB:X03202; NID:G190481; PIDN:AAA60183.1; PID:G190482  
J. Biol. Chem. 261, 6712-6718, 1986  
R:Kim, H.S.; Maeda, N.  
A:Title: Structures of two HaerII-type genes in the human salivary proline-rich proteir  
A:Reference number: A57868; MUID:86196106; PMID:3009472  
A:Accession: B57868  
A:Molecule type: DNA  
A:Residues: 1-166 <KIM>  
A:Cross-references: GB:M13058; NID:G190513; PIDN:AAA98808.1; PID:G190514  
R:Wong, R.S.C.; Bennick, A.  
J. Biol. Chem. 255, 5943-5948, 1980  
A:Title: The primary structure of a salivary calcium-binding proline-rich phosphoprotei  
A:Reference number: A92277; MUID:80204368; PMID:7380845  
A:Contents: Protein C  
A:Accession: A92277  
A:Molecule type: protein  
A:Residues: 17-46 <SCH>  
R:Kim, H.S.; Maeda, N.  
J. Biol. Chem. 261, 6712-6718, 1986  
A:Title: Structures of two HaerII-type genes in the human salivary proline-rich proteir  
A:Reference number: A57868; MUID:86196106; PMID:3009472  
A:Accession: B57868  
A:Molecule type: DNA  
A:Residues: 1-166 <KIM>  
A:Cross-references: GB:M13058; NID:G190513; PIDN:AAA98808.1; PID:G190514  
R:Wong, R.S.C.; Bennick, A.  
J. Biol. Chem. 255, 5943-5948, 1980  
A:Title: The primary structure of a salivary calcium-binding proline-rich phosphoprotei  
A:Reference number: A92277; MUID:80204368; PMID:7380845  
A:Contents: Protein C  
A:Accession: A92277  
A:Molecule type: protein  
A:Residues: 17-19, 'N', '21-166 <WON>  
A:Title: The amino-terminal 46 residues are involved with inhibiting hydroxyapatite form  
A:Wong, R.S.C.; Hofmann, T.; Bennick, A.  
J. Biol. Chem. 254, 4800-4808, 1979  
A:Reference number: A92254; MUID:79173237; PMID:438215  
A:Contents: protein A  
A:Accession: A92254  
A:Molecule type: protein  
A:Residues: 17-19, 'N', '21-122 <WC2>  
R:Schlesinger, D.H.; Hay, D.I.  
in: Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gross  
A:Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent  
A:Reference number: A94425

A;Accession: A94425  
 A;Molecule type: protein  
 A;Residues: 17-122 <SC2>  
 A;Note: the authors call this protein PRP-4  
 R;Semura, S.; Saitoh, E.; Sanada, K.  
 J. Biochem. 87, 1071-1077, 1990  
 A;Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to the human salivary proline-rich proteins (PRP)  
 A;Reference number: A91954; MUID:80227634; PMID:7390979  
 A;Contents: peptide P-C  
 A;Accession: A91954  
 A;Molecule type: protein  
 A;Residues: 123-166 <ICE>  
 R;Hay, D.J.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A;Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A;Reference number: S02562; MUID:89061650; PMID:3196309  
 A;Accession: S02564  
 A;Molecule type: protein  
 A;Residues: 17-166 <HAY>  
 A;Accession: S02563  
 A;Molecule type: protein  
 A;Residues: 47-71 <HAY>  
 R;Schlesinger, D.H.; Hay, D.I.  
 Int. J. Pept. Protein Res. 27, 373-379, 1986  
 A;Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of the human salivary proline-rich proteins from human parotid saliva: relationships of the covalent structure with the primary structure  
 A;Reference number: A98355; MUID:91190884; PMID:1849422  
 A;Accession: G38355  
 A;Status: preliminary  
 A;Molecule type: protein  
 A;Residues: 123-166 <KAU>  
 R;Robinson, R.; Kauffman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A;Title: Primary structure and possible origin of the non-glycosylated basic proline-rich protein from human parotid saliva: relationships of the covalent structure with the primary structure  
 A;Reference number: S06153; MUID:90088384; PMID:2688632  
 A;Accession: S06153  
 A;Molecule type: protein  
 A;Residues: 123-166 <ROB>  
 A;Comment: The proposed biological functions are a highly potent inhibitor of crystal growth  
 C;Genetics:  
 A;Gene: GDB:PRH2  
 A;Cross-references: GDB:119516; OMIM:168790  
 A;Map position: 12p13.2-12p13.2  
 A;Introns: 22/1; 34/1  
 C;Superfamily: proline-rich protein  
 C;Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva  
 F;1-16/Domain: signal sequence #status predicted <SIG>  
 F;17-166/Product: protein C #status experimental <PRC>  
 F;17-122/Product: protein A #status experimental <PRA>  
 F;17-46/Region: apatitic mineral binding  
 F;17-71/Product: PRP-3 #status experimental <PRP3>  
 F;123-166/Product: peptide P-C #status experimental <PPC>  
 F;17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
 F;24,38/Binding site: phosphate (Ser) (covalent) #status experimental  
 Query Match 100.0%; Score 53; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. NO. 0.18;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RGRPGGPPQ 9  
 Db 122 RGRPGGPPQ 130  
 RESULT 2  
 B25372  
 proline-rich phosphoprotein (gene PRH1, Db allele) - human  
 N;Alternate names: salivary acidic proline-rich protein  
 C;Species: Homo sapiens (man)  
 salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human  
 C;Species: Homo sapiens (man)  
 C;Date: 29-Aug-1987 #sequence, revision 29-Aug-1987 #text change 20-Aug-1999  
 C;Accession: B25372; A57868; S02562; G38355; S06153; B27307  
 R;Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 11123-11130, 1985  
 A;Title: Differential RNA splicing and post-translational cleavages in the human salivary proline-rich phosphoprotein precursor PRH1 (allele PIF)  
 A;Reference number: A92492; MUID:85289325; PMID:2993301  
 A;Accession: B25372  
 A;Molecule type: mRNA  
 A;Residues: 1-166 <MAE>  
 A;Cross-references: GB:K03203; NID:G190483; PIDN:AAA60184.1; PID:G190484  
 R;Kim, H.S.; Maeda, N.  
 J. Biol. Chem. 261, 6712-6718, 1986  
 A;Title: Structures of two HaellI-type genes in the human salivary proline-rich protein precursor PRH1 (allele PIF)  
 A;Reference number: A57868; MUID:86196106; PMID:3009472  
 A;Accession: A57868  
 A;Molecule type: DNA  
 A;Residues: 1-166 <KIM>  
 A;Cross-references: GB:M13057; NID:G190511; PIDN:AAA98807.1; PID:G190512  
 R;Hay, D.J.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A;Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A;Reference number: S02562; MUID:89061650; PMID:3196309  
 A;Accession: S02562  
 A;Molecule type: protein  
 A;Residues: 47-71 <HAY>  
 R;Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure with the primary structure  
 A;Reference number: A98355; MUID:91190884; PMID:1849422  
 A;Accession: G38355  
 A;Molecule type: protein  
 A;Residues: 123-166 <KAU>  
 R;Robinson, R.; Kauffman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A;Title: Primary structure and possible origin of the non-glycosylated basic proline-rich protein from human parotid saliva: relationships of the covalent structure with the primary structure  
 A;Reference number: S06153; MUID:90088384; PMID:2688632  
 A;Accession: S06153  
 A;Molecule type: protein  
 A;Residues: 123-166 <ROB>  
 R;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A;Title: Alleles at the PRH1 locus coding for the human salivary acidic proline-rich protein precursor PRH1 (allele PIF)  
 A;Reference number: A27307; MUID:86074309; PMID:3687941  
 A;Contents: allele PIF  
 A;Accession: B27307  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 17-41, 'L', 43-118, 'C', 120-166 <AZE>  
 A;Cross-references: EMBL:K03203  
 C;Genetics:  
 A;Gene: GDB:PRH1  
 A;Cross-references: GDB:119515; OMIM:168730  
 A;Map position: 12p13.2-12p13.2  
 A;Introns: 22/1; 34/1  
 C;Superfamily: proline-rich protein  
 C;Keywords: phosphoprotein; saliva; tandem repeat  
 Query Match 100.0%; Score 53; DB 2; Length 166;  
 Best Local Similarity 100.0%; Pred. NO. 0.18;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RGRPGGPPQ 9  
 Db 122 RGRPGGPPQ 130  
 RESULT 3  
 A27307  
 proline-rich phosphoprotein (gene PRH1, Db allele) - human  
 N;Alternate names: salivary acidic proline-rich protein  
 C;Species: Homo sapiens (man)

C>Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 29-Aug-1997  
 C/Accession: A27307  
 R/Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A/Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
 A/Reference number: A27307; MUID:88074309; PMID:3687941  
 A/Accession: A27307  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: DNA  
 A/Residues: 1-171 <AZE>  
 A/Cross-references: EMBL:K03203  
 C/Genetics:  
 C/Gene: GDB:PRH1  
 A/Cross-references: GDB:119515; OMIM:168730  
 A/Map position: 12p13.2-12p13.2  
 C/Superfamily: proline-rich protein  
 C/Keywords: phosphoprotein

Query Match 100.0%; Score 53; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 0.18; Indels 0; Gaps 0;  
 Matches 9; Conservative 0; Mismatches 0

QY 1 RGRPGPPQ 9  
 |||||  
 Db 127 RGRPGPPQ 135

RESULT 4  
 D38355  
 basic proline-rich peptide IB-8a - human (fragments)  
 C/Species: Homo sapiens (man)  
 C/Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 12-Apr-1995  
 C/Accession: D38355  
 R/Kaufman, D.L.; Bannick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the con-  
 A/Reference number: A38355; MUID:91190884; PMID:1849422  
 A/Accession: D38355  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 1-128 <KAU>  
 C/Superfamily: proline-rich protein

Query Match 84.9%; Score 45; DB 2; Length 128;  
 Best Local Similarity 87.5%; Pred. No. 2.6; Indels 0; Gaps 0;  
 Matches 7; Conservative 1; Mismatches 0

QY 2 GRPGPPQ 9  
 |||||  
 Db 99 GRPGPPQ 106

RESULT 5  
 JH0481  
 basic proline-rich protein Mnp4 - crab-eating macaque  
 C/Species: Macaca fascicularis (crab-eating macaque)  
 C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Aug-1999  
 C/Accession: JH0481; JH0482  
 R/Lin, H.H.; Kousvelari, E.E.; Ann, D.K.  
 Gene 104, 219-226, 1991  
 A/Title: Sequence and expression of the Mnp4 gene encoding basic proline-rich protein in  
 A/Reference number: JH0481; MUID:92009216; PMID:1916292  
 A/Accession: JH0481  
 A/Molecule type: DNA  
 A/Residues: 1-188 <LIN>  
 A/Cross-references: GB:M81322; GB:M61736; NID:G342284; PIDN:AAA36905.1; PID:G942285  
 A/Experimental source: salivary gland  
 A/Accession: JH0482  
 A/Molecule type: mRNA  
 A/Residues: 1-188 <LIN>  
 A/Cross-references: GB:M81321; GB:M61735; NID:G342282; PIDN:AAA36904.1; PID:G342283  
 C/Comment: This protein is a major constituent of both parotid gland and submandibular gland  
 C/Genetics:

A/Introns: 22/1; 34/1; 187/2  
 C/Superfamily: proline-rich protein

Query Match 84.9%; Score 45; DB 2; Length 188;  
 Best Local Similarity 87.5%; Pred. No. 3.6; Indels 0; Gaps 0;  
 Matches 7; Conservative 1; Mismatches 0

QY 2 GRPGPPQ 9  
 |||||  
 Db 95 GRPGPPQ 102

RESULT 6  
 P1HUPF  
 salivary proline-rich glycoprotein precursor PRB2 [validated] - human (fragment)  
 N/Alternate names: basic proline-rich peptide IB-8c precursor; proline-rich protein (c)  
 N/Contains: basic proline-rich peptide IB-4; basic proline-rich peptide P-F  
 C/Species: Homo sapiens (man)  
 C/Date: 15-Nov-1984 #sequence\_revision 12-Apr-1996 #text\_change 08-Dec-2000  
 C/Accession: E25372; A60827; A03294; B38355; A38355; F38355  
 R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 11123-11130, 1985  
 A/Title: Differential RNA splicing and post-translational cleavages in the human saliv-  
 A/Reference number: A92492; MUID:85289325; PMID:2993301  
 A/Accession: E25372  
 A/Molecule type: mRNA  
 A/Residues: 1-251 <MAE>  
 A/Cross-references: GB:K03208; NID:G190509; PIDN:AAA60189.1; PID:G190510  
 R/Mamula, P.W.; Morley, D.J.; Larsen, S.H.; Karn, R.C.  
 Biochem. Genet. 26, 165-175, 1988  
 A/Title: Expression of human salivary protein genes.  
 A/Reference number: A60827; MUID:88240287; PMID:3288192  
 A/Accession: A60827  
 A/Status: not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 175-251 <MAV>  
 R/Saitoh, E.; Isemura, S.; Sanada, K.  
 J. Biochem. 93, 883-888, 1983  
 A/Title: Complete amino acid sequence of a basic proline-rich peptide, P-F, from human  
 A/Reference number: A03294; MUID:83265674; PMID:6874669  
 A/Accession: A03294  
 A/Molecule type: protein  
 A/Residues: 134-194 <SAI>  
 A/Experimental source: saliva  
 R/Kaufman, D.L.; Bannick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the co  
 A/Reference number: A38355; MUID:91190884; PMID:1849422  
 A/Accession: B38355  
 A/Molecule type: protein  
 A/Residues: 134-194 <KAU>  
 A/Experimental source: saliva  
 A/Note: this peptide, which is closely related to that of peptide P-E, contains three  
 A/Accession: A38355  
 A/Molecule type: protein  
 A/Residues: 10-67, 'R' <KA2>  
 A/Accession: F38355  
 A/Molecule type: protein  
 A/Residues: 196-251 <KA3>  
 C/Genetics:  
 C/Gene: GDB:PRB2  
 A/Cross-references: GDB:119512; OMIM:168810  
 A/Map position: 12p13.2-12p13.2  
 C/Superfamily: proline-rich protein  
 C/Keywords: glycoprotein; saliva; tandem repeat  
 F/134-194/Product: basic proline-rich peptide P-F #status experimental <MAT1>  
 F/196-251/Product: basic proline-rich peptide IB-4 #status experimental <MAT2>  
 F/3,65,107/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.9%; Score 45; DB 1; Length 251;  
 Best Local Similarity 87.5%; Pred. No. 4.7; Indels 0; Gaps 0;  
 Matches 7; Conservative 1; Mismatches 0



Qy 2 GRPQGPQ 9  
Db 199 GKQGPQ 206

RESULT 7  
PIHUB6  
salivary proline-rich glycoprotein precursor PRB4 (large allele) (validated) - human  
N/Contains: basic proline-rich protein IB-5; proline-rich peptide P-D  
C/Species: Homo sapiens (man)  
C/Date: 19-Feb-1984 #sequence\_revision 12-Apr-1996 #text\_change 08-Dec-2000  
C/Accession: S03176; S03175; S10890; S38355; A03295; A61294; S62891  
R/Lyons, K.M.; Stein, J.H.; Smithies, O.  
Genetics 120, 267-278, 1988  
A/Title: Length polymorphisms in human proline-rich protein genes generated by intragenic recombination  
A/Reference number: S02127; MUID:89121440; PMID:2851479  
A/Accession: S03176  
A/Status: translation not shown  
A/Molecule type: DNA  
A/Residues: 35-310 <LY1>  
A/Cross-references: EMBL:X07715  
A/Note: large allele  
A/Accession: S03175  
A/Status: translation not shown  
A/Molecule type: DNA  
A/Residues: 35-36, 'E', 38-112, 155-310 <LY2>  
A/Cross-references: EMBL:X07704  
A/Note: medium allele  
A/Accession: S10890  
A/Status: preliminary; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-38, 60-112, 'T', 114-115, 'P', 117-121, 185-271, 'A', 273-310 <LY3>  
A/Cross-references: EMBL:X07882; NID:G95647; PIDN:CAA30729.1; PID:G996670  
R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
J. Biol. Chem. 260, 11123-11130, 1985  
A/Title: Differential RNA splicing and post-translational cleavages in the human salivary proline-rich glycoprotein precursor PRB4 (large allele) (validated)  
A/Reference number: A92492; MUID:85289325; PMID:2993301  
A/Accession: D25372  
A/Molecule type: mRNA  
A/Residues: 1-36, 'E', 38-112, 'T', 114-115, 'P', 117-121, 185-271, 'A', 273-310 <MAE>  
R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
Biochemistry 30, 3351-3356, 1991  
A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the covariant amino acid sequences to the proline-rich region  
A/Reference number: A38355; MUID:91190884; PMID:1849422  
A/Accession: E38355  
A/Molecule type: protein  
A/Residues: 241-254, 'KN', 257-310 <KAU>  
R/Saitoh, E.; Isemura, S.; Sanada, K.  
J. Biochem. 93, 495-502, 1983  
A/Title: Complete amino acid sequence of a basic proline-rich peptide, P-D, from human parotid saliva  
A/Reference number: A03295; MUID:83186122; PMID:6841349  
A/Accession: A03295  
A/Molecule type: protein  
A/Residues: 241-310 <SAI>  
R/Shimomura, H.; Kanai, Y.; Sanada, K.  
J. Biochem. 93, 857-863, 1983  
A/Title: Amino acid sequences of glycopeptides obtained from basic proline-rich glycoprotein precursor PRB4 (large allele) (validated)  
A/Reference number: A61294; MUID:8365671; PMID:6874667  
A/Accession: A61294  
A/Molecule type: protein  
A/Residues: 54-57, 'E', 59-73, 'R', 82-101 <SHI>  
R/Charlton, A.J.; Baxter, N.J.; Lilley, T.H.; Haslam, E.; McDonald, C.J.; Williamson, M.  
FEBS Lett. 382, 289-292, 1996  
A/Title: Tannin interactions with a full-length human salivary proline-rich protein disulfide isomerase  
A/Reference number: S62891; MUID:96184506; PMID:8605987  
A/Accession: S62891  
A/Molecule type: protein  
A/Residues: 241-252 <CHA>  
A/Note: amino end of peptide designated basic proline-rich protein IB-5  
A/Note: it is unclear from the peptide sequence whether this is a product of the PRB2 (H) or PRB4 (L) gene  
C/Genetics:  
A/Gene: GDB:PRB4  
A/Cross-references: GDB:119514; OMIM:180990

A/Map position: 12p13.2-12p13.2  
A/Introns: 22/1; 34/1  
A/Note: the list of introns may be incomplete  
C/Suprafamily: proline-rich protein  
C/Keywords: glycoprotein; saliva; tandem repeat  
F/1-16/Domain: signal sequence #status predicted <SIG>  
F/241-310/Product: proline-rich peptide P-D #status experimental <MAT>  
F/66-87, 171/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F/108, 150, 192, 213, 234/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.9%; Score 45; DB 1; Length 310;  
Best Local Similarity 87.5%; Pred. NO. 5.7;  
Matches 7; Conservative 1; Mismatches 0; Gaps 0;  
Qy 2 GRPQGPQ 9  
Db 244 GKQGPQ 251

RESULT 8  
PIHUB6  
salivary proline-rich phosphoprotein precursor PRB1 (large allele) (validated) - human  
N/Contains: peptide IB-1; peptide P-E (peptide IB-9); peptide P-F; peptide P-H  
C/Species: Homo sapiens (man)  
C/Date: 04-Dec-1986 #sequence\_revision 12-Apr-1996 #text\_change 08-Dec-2000  
C/Accession: B40750; C40750; C25372; S02128; S02127; A03293; A90502; A91974; A91975; A91976; A91977; A91978; A91979; A91980; A91981; A91982; A91983; A91984; A91985; A91986; A91987; A91988; A91989; A91990; A91991; A91992; A91993; A91994; A91995; A91996; A91997; A91998; A91999; A92000; A92001; A92002; A92003; A92004; A92005; A92006; A92007; A92008; A92009; A92010; A92011; A92012; A92013; A92014; A92015; A92016; A92017; A92018; A92019; A92020; A92021; A92022; A92023; A92024; A92025; A92026; A92027; A92028; A92029; A92030; A92031; A92032; A92033; A92034; A92035; A92036; A92037; A92038; A92039; A92040; A92041; A92042; A92043; A92044; A92045; A92046; A92047; A92048; A92049; A92050; A92051; A92052; A92053; A92054; A92055; A92056; A92057; A92058; A92059; A92060; A92061; A92062; A92063; A92064; A92065; A92066; A92067; A92068; A92069; A92070; A92071; A92072; A92073; A92074; A92075; A92076; A92077; A92078; A92079; A92080; A92081; A92082; A92083; A92084; A92085; A92086; A92087; A92088; A92089; A92090; A92091; A92092; A92093; A92094; A92095; A92096; A92097; A92098; A92099; A92100; A92101; A92102; A92103; A92104; A92105; A92106; A92107; A92108; A92109; A92110; A92111; A92112; A92113; A92114; A92115; A92116; A92117; A92118; A92119; A92120; A92121; A92122; A92123; A92124; A92125; A92126; A92127; A92128; A92129; A92130; A92131; A92132; A92133; A92134; A92135; A92136; A92137; A92138; A92139; A92140; A92141; A92142; A92143; A92144; A92145; A92146; A92147; A92148; A92149; A92150; A92151; A92152; A92153; A92154; A92155; A92156; A92157; A92158; A92159; A92160; A92161; A92162; A92163; A92164; A92165; A92166; A92167; A92168; A92169; A92170; A92171; A92172; A92173; A92174; A92175; A92176; A92177; A92178; A92179; A92180; A92181; A92182; A92183; A92184; A92185; A92186; A92187; A92188; A92189; A92190; A92191; A92192; A92193; A92194; A92195; A92196; A92197; A92198; A92199; A92200; A92201; A92202; A92203; A92204; A92205; A92206; A92207; A92208; A92209; A92210; A92211; A92212; A92213; A92214; A92215; A92216; A92217; A92218; A92219; A92220; A92221; A92222; A92223; A92224; A92225; A92226; A92227; A92228; A92229; A92230; A92231; A92232; A92233; A92234; A92235; A92236; A92237; A92238; A92239; A92240; A92241; A92242; A92243; A92244; A92245; A92246; A92247; A92248; A92249; A92250; A92251; A92252; A92253; A92254; A92255; A92256; A92257; A92258; A92259; A92260; A92261; A92262; A92263; A92264; A92265; A92266; A92267; A92268; A92269; A92270; A92271; A92272; A92273; A92274; A92275; A92276; A92277; A92278; A92279; A92280; A92281; A92282; A92283; A92284; A92285; A92286; A92287; A92288; A92289; A92290; A92291; A92292; A92293; A92294; A92295; A92296; A92297; A92298; A92299; A92300; A92301; A92302; A92303; A92304; A92305; A92306; A92307; A92308; A92309; A92310; A92311; A92312; A92313; A92314; A92315; A92316; A92317; A92318; A92319; A92320; A92321; A92322; A92323; A92324; A92325; A92326; A92327; A92328; A92329; A92330; A92331; A92332; A92333; A92334; A92335; A92336; A92337; A92338; A92339; A92340; A92341; A92342; A92343; A92344; A92345; A92346; A92347; A92348; A92349; A92350; A92351; A92352; A92353; A92354; A92355; A92356; A92357; A92358; A92359; A92360; A92361; A92362; A92363; A92364; A92365; A92366; A92367; A92368; A92369; A92370; A92371; A92372; A92373; A92374; A92375; A92376; A92377; A92378; A92379; A92380; A92381; A92382; A92383; A92384; A92385; A92386; A92387; A92388; A92389; A92390; A92391; A92392; A92393; A92394; A92395; A92396; A92397; A92398; A92399; A92400; A92401; A92402; A92403; A92404; A92405; A92406; A92407; A92408; A92409; A92410; A92411; A92412; A92413; A92414; A92415; A92416; A92417; A92418; A92419; A92420; A92421; A92422; A92423; A92424; A92425; A92426; A92427; A92428; A92429; A92430; A92431; A92432; A92433; A92434; A92435; A92436; A92437; A92438; A92439; A92440; A92441; A92442; A92443; A92444; A92445; A92446; A92447; A92448; A92449; A92450; A92451; A92452; A92453; A92454; A92455; A92456; A92457; A92458; A92459; A92460; A92461; A92462; A92463; A92464; A92465; A92466; A92467; A92468; A92469; A92470; A92471; A92472; A92473; A92474; A92475; A92476; A92477; A92478; A92479; A92480; A92481; A92482; A92483; A92484; A92485; A92486; A92487; A92488; A92489; A92490; A92491; A92492; A92493; A92494; A92495; A92496; A92497; A92498; A92499; A92500; A92501; A92502; A92503; A92504; A92505; A92506; A92507; A92508; A92509; A92510; A92511; A92512; A92513; A92514; A92515; A92516; A92517; A92518; A92519; A92520; A92521; A92522; A92523; A92524; A92525; A92526; A92527; A92528; A92529; A92530; A92531; A92532; A92533; A92534; A92535; A92536; A92537; A92538; A92539; A92540; A92541; A92542; A92543; A92544; A92545; A92546; A92547; A92548; A92549; A92550; A92551; A92552; A92553; A92554; A92555; A92556; A92557; A92558; A92559; A92560; A92561; A92562; A92563; A92564; A92565; A92566; A92567; A92568; A92569; A92570; A92571; A92572; A92573; A92574; A92575; A92576; A92577; A92578; A92579; A92580; A92581; A92582; A92583; A92584; A92585; A92586; A92587; A92588; A92589; A92590; A92591; A92592; A92593; A92594; A92595; A92596; A92597; A92598; A92599; A92600; A92601; A92602; A92603; A92604; A92605; A92606; A92607; A92608; A92609; A92610; A92611; A92612; A92613; A92614; A92615; A92616; A92617; A92618; A92619; A92620; A92621; A92622; A92623; A92624; A92625; A92626; A92627; A92628; A92629; A92630; A92631; A92632; A92633; A92634; A92635; A92636; A92637; A92638; A92639; A92640; A92641; A92642; A92643; A92644; A92645; A92646; A92647; A92648; A92649; A92650; A92651; A92652; A92653; A92654; A92655; A92656; A92657; A92658; A92659; A92660; A92661; A92662; A92663; A92664; A92665; A92666; A92667; A92668; A92669; A92670; A92671; A92672; A92673; A92674; A92675; A92676; A92677; A92678; A92679; A92680; A92681; A92682; A92683; A92684; A92685; A92686; A92687; A92688; A92689; A92690; A92691; A92692; A92693; A92694; A92695; A92696; A92697; A92698; A92699; A92700; A92701; A92702; A92703; A92704; A92705; A92706; A92707; A92708; A92709; A92710; A92711; A92712; A92713; A92714; A92715; A92716; A92717; A92718; A92719; A92720; A92721; A92722; A92723; A92724; A92725; A92726; A92727; A92728; A92729; A92730; A92731; A92732; A92733; A92734; A92735; A92736; A92737; A92738; A92739; A92740; A92741; A92742; A92743; A92744; A92745; A92746; A92747; A92748; A92749; A92750; A92751; A92752; A92753; A92754; A92755; A92756; A92757; A92758; A92759; A92760; A92761; A92762; A92763; A92764; A92765; A92766; A92767; A92768; A92769; A92770; A92771; A92772; A92773; A92774; A92775; A92776; A92777; A92778; A92779; A92780; A92781; A92782; A92783; A92784; A92785; A92786; A92787; A92788; A92789; A92790; A92791; A92792; A92793; A92794; A92795; A92796; A92797; A92798; A92799; A92800; A92801; A92802; A92803; A92804; A92805; A92806; A92807; A92808; A92809; A92810; A92811; A92812; A92813; A92814; A92815; A92816; A92817; A92818; A92819; A92820; A92821; A92822; A92823; A92824; A92825; A92826; A92827; A92828; A92829; A92830; A92831; A92832; A92833; A92834; A92835; A92836; A92837; A92838; A92839; A92840; A92841; A92842; A92843; A92844; A92845; A92846; A92847; A92848; A92849; A92850; A92851; A92852; A92853; A92854; A92855; A92856; A92857; A92858; A92859; A92860; A92861; A92862; A92863; A92864; A92865; A92866; A92867; A92868; A92869; A92870; A92871; A92872; A92873; A92874; A92875; A92876; A92877; A92878; A92879; A92880; A92881; A92882; A92883; A92884; A92885; A92886; A92887; A92888; A92889; A92890; A92891; A92892; A92893; A92894; A92895; A92896; A92897; A92898; A92899; A92900; A92901; A92902; A92903; A92904; A92905; A92906; A92907; A92908; A92909; A92910; A92911; A92912; A92913; A92914; A92915; A92916; A92917; A92918; A92919; A92920; A92921; A92922; A92923; A92924; A92925; A92926; A92927; A92928; A92929; A92930; A92931; A92932; A92933; A92934; A92935; A92936; A92937; A92938; A92939; A92940; A92941; A92942; A92943; A92944; A92945; A92946; A92947; A92948; A92949; A92950; A92951; A92952; A92953; A92954; A92955; A92956; A92957; A92958; A92959; A92960; A92961; A92962; A92963; A92964; A92965; A92966; A92967; A92968; A92969; A92970; A92971; A92972; A92973; A92974; A92975; A92976; A92977; A92978; A92979; A92980; A92981; A92982; A92983; A92984; A92985; A92986; A92987; A92988; A92989; A92990; A92991; A92992; A92993; A92994; A92995; A92996; A92997; A92998; A92999; A93000; A93001; A93002; A93003; A93004; A93005; A93006; A93007; A93008; A93009; A93010; A93011; A93012; A93013; A93014; A93015; A93016; A93017; A93018; A93019; A93020; A93021; A93022; A93023; A93024; A93025; A93026; A93027; A93028; A93029; A93030; A93031; A93032; A93033; A93034; A93035; A93036; A93037; A93038; A93039; A93040; A93041; A93042; A93043; A93044; A93045; A93046; A93047; A93048; A93049; A93050; A93051; A93052; A93053; A93054; A93055; A93056; A93057; A93058; A93059; A93060; A93061; A93062; A93063; A93064; A93065; A93066; A93067; A93068; A93069; A93070; A93071; A93072; A93073; A93074; A93075; A93076; A93077; A93078; A93079; A93080; A93081; A93082; A93083; A93084; A93085; A93086; A93087; A93088; A93089; A93090; A93091; A93092; A93093; A93094; A93095; A93096; A93097; A93098; A93099; A93100; A93101; A93102; A93103; A93104; A93105; A93106; A93107; A93108; A93109; A93110; A93111; A93112; A93113; A93114; A93115; A93116; A93117; A93118; A93119; A93120; A93121; A93122; A93123; A93124; A93125; A93126; A93127; A93128; A93129; A93130; A93131; A93132; A93133; A93134; A93135; A93136; A93137; A93138; A93139; A93140; A93141; A93142; A93143; A93144; A93145; A93146; A93147; A93148; A93149; A93150; A93151; A93152; A93153; A93154; A93155; A93156; A93157; A93158; A93159; A93160; A93161; A93162; A93163; A93164; A93165; A93166; A93167; A93168; A93169; A93170; A93171; A93172; A93173; A93174; A93175; A93176; A93177; A93178; A93179; A93180; A93181; A93182; A93183; A93184; A93185; A93186; A93187; A93188; A93189; A93190; A93191; A93192; A93193; A93194; A93195; A93196; A93197; A93198; A93199; A93200; A93201; A93202; A93203; A93204; A93205; A93206; A93207; A93208; A93209; A93210; A93211; A93212; A93213; A93214; A93215; A93216; A93217; A93218; A93219; A93220; A93221; A93222; A93223; A93224; A93225; A93226; A93227; A93228; A93229; A93230; A93231; A93232; A93233; A93234; A93235; A93236; A93237; A93238; A93239; A93240; A93241; A93242; A93243; A93244; A93245; A93246; A93247; A93248; A93249; A93250; A93251; A93252; A93253; A93254; A93255; A93256; A93257; A93258; A93259; A93260; A93261; A93262; A93263; A93264; A93265; A93266; A93267; A93268; A93269; A93270; A93271; A93272; A93273; A93274; A93275; A93276; A93277; A93278; A93279; A93280; A93281; A93282; A93283; A93284; A93285; A93286; A93287; A93288; A93289; A93290; A93291; A93292; A93293; A93294; A93295; A93296; A93297; A93298; A93299; A93300; A93301; A93302; A93303; A93304; A93305; A93306; A93307; A93308; A93309; A93310; A93311; A93312; A93313; A93314; A93315; A93316; A93317; A93318; A93319; A93320; A93321; A93322; A93323; A93324; A93325; A93326; A93327; A93328; A93329; A93330; A93331; A93332; A93333; A93334; A93335; A93336; A93337; A93338; A93339; A93340; A93341; A93342; A93343; A93344; A93345; A93346; A93347; A93348; A93349; A93350; A93351; A93352; A93353; A93354; A93355; A93356; A93357; A93358; A93359; A93360; A93361; A93362; A93363; A93364; A93365; A93366; A93367; A93368; A93369; A93370; A93371; A93372; A93373; A93374; A93375; A93376; A93377; A93378; A93379; A93380; A93381; A93382; A93383; A93384; A93385; A93386; A93387; A93388; A93389; A93390; A93391; A93392; A93393; A93394; A93395; A93396; A93397; A93398; A93399; A93400; A93401; A93402; A93403; A93404; A93405; A93406; A93407; A93408; A93409; A93410; A93411; A93412; A93413; A93414; A93415; A93416; A93417; A93418; A93419; A93420; A93

```

A/Accession: A03293
A/Molecule type: protein
A/Residues: 17-38, 'AP', 41-51, 92-148, 'R', 150-152 <KA2>
A/Note: among nine basic proline-rich peptides isolated from the saliva, this peptide is
A/Accession: A90502
A/Molecule type: protein
A/Residues: 275-336, 'S', 338-392 <KAU>
R/Saitoh, E.; Isemura, S.; Sanada, K.
J. Biochem. 94, 1991-1999, 1983
A/Title: Further fractionation of basic proline-rich peptides from human parotid saliva
A/Reference number: A91974; MUID: 84161824; PMID: 6671974
A/Contents: P-H
A/Accession: A91974
A/Molecule type: protein
A/Residues: 'S', 338-392 <SAI>
R/Azen, E.; Lyons, K.M.; McDonigal, T.; Barrett, N.L.; Clements, L.S.; Maeda, N.; Vanin,
Proc. Natl. Acad. Sci. U.S.A. 81, 5561-5565, 1984
A/Reference number: A94005; MUID: 84298176; PMID: 6089212
A/Accession: A05261
A/Molecule type: DNA
A/Residues: 35-39, 'P', 41-84, 'G', 86, 'R', 87-154, 'R', 218-246; 300-306, 'T', 308-329, 'C', 331-38
A/Accession: A05262
A/Molecule type: DNA
A/Residues: 'N', 57-59, 'A', 61-69; 334-336, 'S', 338-339, 'R', 341-392 <AZ5>
R/Kaufman, D.; Wong, R.; Bennick, A.; Keller, P.
Biochemistry 21, 6558-6562, 1982
A/Title: Basic proline-rich proteins from human parotid saliva: complete covalent struct
A/Reference number: A90464; MUID: 83101329; PMID: 6924859
A/Contents: IB-9
A/Accession: A90464
A/Molecule type: protein
A/Residues: 92-127, 'R', 129-148, 'R', 150-152 <KA3>
R/Isemura, S.; Saitoh, E.; Sanada, K.
J. Biochem. 91, 2067-2075, 1982
A/Title: Fractionation and characterization of basic proline-rich peptides of human par
A/Reference number: A91966; MUID: 83007119; PMID: 7118863
A/Contents: P-E
A/Accession: A91966
A/Molecule type: protein
A/Residues: 92-127, 'R', 129-148, 'R', 150-152 <ISB>
C/Comment: This peptide contains 21-residue repeats, two of which have internal 7-residu
C/Genetics:
A/Gene: GDB:PRB1
A/Cross-references: GDB:119511; OMIM:180989
A/Map position: 12p13.2-12p13.2
A/Note: each of the tandem repeats contains a candidate splice acceptor site, and severa
C/Superfamily: proline-rich protein
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-51/Domain: duplication; parotid gland; phosphoprotein; pyroglutam
F:51-92-152/Product: basic proline-rich peptide IB-1 #status experimental <IB1>
F:92-152/Product: basic proline-rich peptide P-E #status experimental <PPE>
F:275-332/Product: basic proline-rich peptide IS-6 #status experimental <PIB6>
F:275-335/Product: basic proline-rich peptide P-F #status experimental <PPF>
F:337-392/Product: basic proline-rich peptide P-H #status experimental <PPH>
F:17/Modified site: pyrrolidone carboxylic acid (gln) (in mature form) #status experimen
F:24/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 84.9%; Score 45; DB 1; Length 392;
Best Local Similarity 87.5%; Pred. No. 7.1;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGPQ 9
Db 238 GRPQGPQ 245

RESULT 9
D29149
Proline-rich protein - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 20-Aug-1999
C/Accession: D29149
R/Clements, S.; Mehansho, H.; Carlson, D.M.

J. Biol. Chem. 260, 13471-13477, 1985
A/Title: Novel multigene families encoding highly repetitive peptide sequences. Sequen
A/Reference number: A92501; MUID: 86033799; PMID: 3840480
A/Contents: Clone pUMP4
A/Accession: D29149
A/Molecule type: mRNA
A/Residues: 1-188 <CLE>
A/Cross-references: GB:U19419; NID: g200542; PIDN: AAA40002.1; PID: g200543
C/Superfamily: proline-rich protein

Query Match 81.1%; Score 43; DB 2; Length 188;
Best Local Similarity 87.5%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GRPQGPQ 9
Db 174 GRPQGPQ 181

RESULT 10
B36298
Proline-rich protein PRB3S (cys) - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 29-Aug-1997
C/Accession: B36298
R/Azen, E.A.; Minaguchi, K.; Latreille, P.; Kim, H.S.
Am. J. Hum. Genet. 47, 686-697, 1990
A/Title: Alleles at the PRB3 locus coding for a disulfide-bonded human salivary prolin
A/Reference number: A36298; MUID: 91022705; PMID: 2171329
A/Accession: B36298
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-212 <AZE>
C/Genetics:
A/Gene: GDB:PRB3
A/Cross-references: GDB:119513; OMIM:168840
A/Map position: 12p13.2-12p13.2
C/Superfamily: proline-rich protein

Query Match 81.1%; Score 43; DB 2; Length 212;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGPQ 8
Db 167 GRPQGPQ 173

RESULT 11
B48013
proline-rich proteoglycan 2 precursor, parotid - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 20-Aug-1999
C/Accession: B48013
R/Castle, A.M.; Castle, J.D.
J. Biol. Chem. 268, 20490-20496, 1993
A/Title: Novel secretory proline-rich proteoglycans from rat parotid. Cloning and char
A/Reference number: A48013; MUID: 93388626; PMID: 8376404
A/Accession: B48013
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-295 <CAS>
A/Cross-references: GB:U17318; NID: g310199; PIDN: AAA03074.1; PID: g310200
C/Superfamily: proline-rich protein
C/Keywords: extracellular protein; glycoprotein; tandem repeat

Query Match 81.1%; Score 43; DB 2; Length 295;
Best Local Similarity 87.5%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GRPQGPQ 9
Db 273 GRPQGPQ 280

```

## RESULT 12

A28996

Proline-rich protein M14 precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 20-Aug-1999  
 C:Accession: A28996  
 R:Ann, D.K.; Smith, K.; Carlson, D.M.  
 J. Biol. Chem. 263, 10887-10893, 1988  
 A:Title: Molecular evolution of the mouse proline-rich protein multigene family. Insertion  
 A:Reference number: A28996; MUID:88273214; PMID:2839509  
 A:Accession: A28996  
 A:Molecule type: DNA  
 A:Residues: 1-317 <NN>  
 A:Cross-references: GB:M3236; GB:J03891; NID:G200535; PIDN:AAA53048.1; PID:G567232  
 C:Genetics:

A:Introns: 22/1

C:Superfamily: proline-rich protein

C:Keywords: saliva

F:1-15/Domain: signal sequence #status predicted &lt;SIG&gt;

F:16-317/Product: proline-rich protein M14 #status predicted &lt;MAT&gt;

Query Match 81.1%; Score 43; DB 2; Length 317;

Best Local Similarity 87.5%; Pred. No. 12;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPQ 9

Db 303 GNPQGPQ 310

RESULT 13

A48013

Proline-rich proteoglycan 1 precursor, parotid - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 02-Jun-1995 #sequence\_revision 02-Jun-1995 #text\_change 20-Aug-1999  
 C:Accession: A48013  
 R:Castle, A.M.; Castle, J.D.  
 J. Biol. Chem. 268, 20490-20496, 1993  
 A:Title: Novel secretory proline-rich proteoglycans from rat parotid. Cloning and charac  
 A:Reference number: A48013; MUID:93388626; PMID:8376404  
 A:Accession: A48013  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-170 <CAS>  
 A:Cross-references: GB:L17317; NID:G310197; PIDN:AAA03073.1; PID:G310198  
 C:Superfamily: proline-rich protein  
 C:Keywords: extracellular protein; glycoprotein; tandem repeat

Query Match 79.2%; Score 42; DB 2; Length 170;

Best Local Similarity 77.8%; Pred. No. 9.9;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRPQGPQ 9

Db 119 QGPGGPQ 127

RESULT 14

PIRT3

acidic proline-rich protein precursor - rat  
 N:Alternate names: PRP  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 22-Jun-1999  
 C:Accession: A03296  
 R:Ziemer, M.A.; Swain, W.F.; Rutter, W.J.; Clements, S.; Ann, D.K.; Carlson, D.M.  
 J. Biol. Chem. 259, 10475-10480, 1984  
 A:Title: Nucleotide sequence analysis of a proline-rich protein cDNA and peptide homolog  
 A:Reference number: A03296; MUID:84289443; PMID:6547951  
 A:Accession: A03296  
 A:Molecule type: mRNA  
 A:Residues: 1-206 <ZIE>

Query Match 79.2%; Score 42; DB 2; Length 170;

Best Local Similarity 77.8%; Pred. No. 40;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRPQGPQ 9

Db 119 QGPGGPQ 127

RESULT 15

S53590

hypothetical protein YCR077c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein YCR078c  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 13-May-1995 #sequence\_revision 01-Dec-1995 #text\_change 29-Oct-1999  
 C:Accession: S53590; S55867; S19491; S19492  
 R:Jimenez, A.  
 submitted to the EMBL Data Library, December 1992  
 A:Reference number: S53589  
 A:Accession: S53590  
 A:Molecule type: DNA  
 A:Residues: 1-797 <JIN>  
 A:Cross-references: EMBL:X59720; NID:G1907116; PIDN:CAA42266.1; PID:G1907  
 A:Note: this is a correction to the sequence in reference S19486  
 R:Ballesta, J.P.G.; Franco, I.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E.  
 submitted to the Protein Sequence Database, March 1992  
 A:Reference number: S19486  
 A:Accession: S19491  
 A:Molecule type: DNA  
 A:Residues: 135-641, 'BH' <BAL>  
 A:Cross-references: EMBL:X59720; MIPS:YCR077c  
 A:Note: this sequence was incomplete due to a frameshift error  
 A:Note: this sequence has been revised in reference S55867  
 A:Accession: S19492  
 A:Molecule type: DNA  
 A:Residues: 'MNN', '64', 'SGRYLLTPPDLHP', '79-80', 'PSSKPEQVYDFKPEH', '83-132', 'RKWP', '137', 'YS'

Query Match 79.2%; Score 42; DB 2; Length 797;

Best Local Similarity 77.8%; Pred. No. 40;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRPQGPQ 9

Db 168 QGLPGGPQ 176

Search completed: April 6, 2004, 16:16:51

Job time: 10.0935 secs

A:Cross-references: GB:K02247; NID:G206395; PIDN:AAA1949.1; PID:G206396  
 C:Comment: This protein contains six 18- to 19-residue repeats.  
 C:Comment: This protein may protect teeth by binding to tannins.  
 C:Superfamily: proline-rich protein  
 C:Keywords: duplication; parotid gland; saliva; tandem repeat  
 F:1-13/Domain: signal sequence #status predicted <SIG>  
 F:14-206/Product: acidic proline-rich protein #status predicted <MAT>  
 F:80-189/Region: 18-residue repeats

Query Match 79.2%; Score 42; DB 1; Length 206;

Best Local Similarity 77.8%; Pred. No. 12;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRPQGPQ 9

Db 119 QGPGGPQ 127

RESULT 15

S53590

hypothetical protein YCR077c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein YCR078c  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 13-May-1995 #sequence\_revision 01-Dec-1995 #text\_change 29-Oct-1999  
 C:Accession: S53590; S55867; S19491; S19492  
 R:Jimenez, A.  
 submitted to the EMBL Data Library, December 1992  
 A:Reference number: S53589  
 A:Accession: S53590  
 A:Molecule type: DNA  
 A:Residues: 1-797 <JIN>  
 A:Cross-references: EMBL:X59720; NID:G1907116; PIDN:CAA42266.1; PID:G1907  
 A:Note: this is a correction to the sequence in reference S19486  
 R:Ballesta, J.P.G.; Franco, I.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E.  
 submitted to the Protein Sequence Database, March 1992  
 A:Reference number: S55867; MUID:95373282; PMID:7645349  
 A:Accession: S55867  
 A:Molecule type: DNA  
 A:Residues: 1-797 <ROD>  
 A:Cross-references: EMBL:X59720; NID:G1907116; PIDN:CAA42266.1; PID:G1907  
 A:Note: this is a correction to the sequence in reference S19486  
 R:Ballesta, J.P.G.; Franco, I.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E.  
 submitted to the Protein Sequence Database, March 1992  
 A:Reference number: S19486  
 A:Accession: S19491  
 A:Molecule type: DNA  
 A:Residues: 135-641, 'BH' <BAL>  
 A:Cross-references: EMBL:X59720; MIPS:YCR077c  
 A:Note: this sequence was incomplete due to a frameshift error  
 A:Note: this sequence has been revised in reference S55867  
 A:Accession: S19492  
 A:Molecule type: DNA  
 A:Residues: 'MNN', '64', 'SGRYLLTPPDLHP', '79-80', 'PSSKPEQVYDFKPEH', '83-132', 'RKWP', '137', 'YS'

Query Match 79.2%; Score 42; DB 2; Length 797;

Best Local Similarity 77.8%; Pred. No. 40;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GRPQGPQ 9

Db 168 QGLPGGPQ 176

Search completed: April 6, 2004, 16:16:51

Job time: 10.0935 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 5.80374 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-5  
Perfect score: 53  
Sequence: 1 RGRPGPPQ 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	166	1 PRPC_HUMAN	P02810 homo sapien
2	45	84.9	234	1 PRPM_HUMAN	P10161 homo sapien
3	45	84.9	247	1 PRB4_HUMAN	P10163 homo sapien
4	45	84.9	251	1 PRP2_HUMAN	P02812 homo sapien
5	45	84.9	276	1 PRPL_HUMAN	P10162 homo sapien
6	45	84.9	331	1 PRP1_HUMAN	P04280 homo sapien
7	42	78.2	124	1 CES4_HUMAN	Q9bxq7 homo sapien
8	42	78.2	206	1 PRP3_RAT	P04474 rattus norv
9	42	78.2	797	1 PAT1_YEAST	P25644 saccharomyc
10	41	77.4	125	1 SVB_IOLPE	P47194 loligo peal
11	41	77.4	1276	1 SCAP_CRIGR	P97260 cricetulus
12	40	75.5	61	1 PRPE_HUMAN	P02811 homo sapien
13	40	75.5	96	1 PRFS_HUMAN	P04281 homo sapien
14	40	75.5	172	1 PRP2_RAT	P10164 rattus norv
15	40	75.5	174	1 PRPP_HUMAN	P81489 homo sapien
16	39	73.6	842	1 ORP7_HUMAN	Q9bzt2 homo sapien
17	38	71.7	11	1 BPP_AKGHP	P04562 agkistrodon
18	38	71.7	296	1 PWP_MOUSE	P05143 mus musculus
19	38	71.7	416	1 P2R2_RAT	P43253 rattus norv
20	38	71.7	551	1 ERF_MOUSE	P70459 mus musculus
21	38	71.7	5262	1 EPFL_HUMAN	P58107 homo sapien
22	38	71.7	5262	1 MLL2_HUMAN	O14686 homo sapien
23	37	69.8	113	1 VAM2_XENLA	P47193 xenopus lae
24	37	69.8	261	1 PRP2_MOUSE	P05142 mus musculus
25	37	69.8	323	1 LSG4_HUMAN	P256470 homo sapien
26	37	69.8	373	1 MGE1_HUMAN	Q9ubf1 homo sapien
27	37	69.8	520	1 GAG_SIVAI	P27972 simian immu
28	37	69.8	740	1 GAG_SMRVH	P21411 squirrel mo
29	37	69.8	1226	1 CDN1_HUMAN	O81wy9 homo sapien
30	37	69.8	1336	1 W146_HUMAN	Q9c018 homo sapien
31	37	69.8	1806	1 CALB_HUMAN	P12107 homo sapien
32	37	69.8	3149	1 TSGU_EBV	P03186 Epstein-Bar
33	36	67.9	209	1 NT5_HUMAN	P34131 rattus norv

RESULT 1  
PRPC\_HUMAN  
ID PRPC\_HUMAN STANDARD; PRT; 166 AA.  
AC P02810;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains: Peptide P-C].  
DE PRH1 AND PRH2.  
GN PRH1 AND PRH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).  
RX MEDLINE=86196106; PubMed=3009472;  
RA Kim H.-S., Maeda N.;  
RT "Structures of two HaeIII-type genes in the human salivary proline-rich protein multigene family."  
RL J. Biol. Chem. 261:6712-6718(1986).  
RN [2]  
RP SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).  
RX MEDLINE=85289325; PubMed=2993301;  
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system."  
RL J. Biol. Chem. 260:11123-11130(1985).  
RN [3]  
RP SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).  
RX MEDLINE=89061650; PubMed=3196309;  
RA Hay D.I., Bennick A., Schlesinger D.H., Mineguchi K., Madapallimattam G., Schluckebier S.K.;  
RT "The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f)."; Biochem. J. 255:15-21(1988).  
RN [4]  
RP SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).  
RX MEDLINE=88074305; PubMed=3687941;  
RA Azen E.A., Kim H.S., Goodman P., Flynn S., Maeda N.;  
RT "Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich proteins Pa, Db, and PIF."  
RL Am. J. Hum. Genet. 41:1035-1047(1987).  
RN [5]  
RP SEQUENCE OF 17-166 (PRP-2).  
RX MEDLINE=86222915; PubMed=3710693;  
RA Schlesinger D.H., Hay D.I.;  
RT "Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva."  
RL Int. J. Rept. Protein Res. 27:373-379(1986).  
RN [6]  
RP SEQUENCE OF 17-166 (PROTEIN C).  
RX MEDLINE=80204368; PubMed=7380845;  
RA Wong R.S.C., Bennick A.;

#### ALIGNMENTS

34	36	67.9	210	1	NT5_HUMAN	P34130 homo sapien
35	36	67.9	243	1	YX42_MYCTU	O53392 mycobacteri
36	36	67.9	293	1	B4G3_HUMAN	O60512 h beta-1,4-
37	36	67.9	548	1	ERF_HUMAN	P50548 homo sapien
38	36	67.9	633	1	ROR_HUMAN	Q43390 homo sapien
39	36	67.9	740	1	CEB1_HUMAN	Q15027 homo sapien
40	36	67.9	753	1	SK30_HUMAN	O94993 homo sapien
41	36	67.9	1009	1	FAK2_MOUSE	O9qvp9 h protein t
42	36	67.9	1278	1	SCAP_HUMAN	Q12770 homo sapien
43	35.5	67.0	664	1	UL47_HSV1P	P08313 herpes simp
44	35.5	67.0	693	1	UL47_HSV1I	P10231 herpes simp
45	35	66.0	59	1	PART_HUMAN	Q9npd0 homo sapien

RT "The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A." J. Biol. Chem. 255:5943-5948 (1980). [7]

RT SEQUENCE OF 17-46 (PROTEIN C).

RT MEDLINE=81191179; PubMed=7228490;

RA Schlesinger D.H., Hay D.I.,

RT "Primary structure of the active tryptic fragments of human and monkey salivary anionic proline-rich proteins." Int. J. Pept. Protein Res. 17:34-41 (1981). [8]

RT SEQUENCE OF 17-122 (PROTEIN A).

RT MEDLINE=79173237; PubMed=438215;

RA Wong R.S.C., Hofmann T., Bennick A.,

RT "The complete primary structure of a proline-rich phosphoprotein from human saliva." J. Biol. Chem. 254:4800-4808 (1979). [9]

RT SEQUENCE OF 17-122 (PROTEIN A).

RA Schlesinger D.H., Hay D.I.;

RT "Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent inhibitor of calcium phosphate precipitation in human parotid saliva." (In) Gross E., Meienhofer J. (eds.); Peptides: structure and biological function (Proceedings of the 6th American peptide symposium), pp.133-136, Pierce Chemical Co., Rockford IL. (1979). [10]

RT SEQUENCE OF 123-166 (PEPTIDE P-C).

RT MEDLINE=80227634; PubMed=7390979;

RA Isemura S., Saichoh E., Sanada K.;

RT "The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to a salivary proline-rich phosphoprotein, protein C." J. Biochem. 87:1071-1077 (1980). [11]

RT VARIANT PRH2-3 LYS-163.

RA Azen E.A.;

RT "A frequent mutation in the acidic proline-rich protein gene, PRH2, causing a Q147K change closely adjacent to the bacterial binding domain of the cognate salivary PRP (Pr1') in Afro-Americans." Hum. Mutat. 12:72-73 (1998).

CC "FUNCTION: PRP's act as highly potent inhibitors of crystal growth of calcium phosphates. They provide a protective and reparative environment for dental enamel which is important for the integrity of the teeth.

CC "SUBCELLULAR LOCATION: Secreted.

CC "PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIP-S yield PRP-4, PRP-3 (protein A), and PIP-F, respectively.

CC "POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele; allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is also known as PR1'.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; K03202; AAA60183.1; -

DR EMBL; K03203; AAA60184.1; -

DR EMBL; M13057; AAA98807.1; -

DR EMBL; M13058; AAA98808.1; -

DR Genew; HGNC:9366; PRH1.

DR Genew; HGNC:9367; PRH2.

DR MIM; 168730; -

DR MIM; 168790; -

DR MIM; 168710; -

DR GO; GO:0005615; C:extracellular space; TAS.

KW Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism; Pyrrolidone carboxylic acid.

FT SIGNAL 1 16  
FT CHAIN 17 166  
FT CHAIN 17 122  
FT CHAIN 17 122  
FT CHAIN 123 166  
FT DOMAIN 17 46  
FT MOD\_RES 17 17  
FT MOD\_RES 24 24  
FT MOD\_RES 38 38  
FT VARIANT 20 20  
FT VARIANT 66 66  
FT VARIANT 163 163  
FT CONFLICT 41 41  
FT SEQUENCE 166 AA; 17017 MW; A7DF62BF94E3C3EF CRC64;

Query Match 100.0%; Score 53; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 0.099;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGPPQ 9  
DB 122 RGRPQGPPQ 130

RESULT 2  
ID PRPM HUMAN STANDARD; PRT; 234 AA.  
AC P10161; P02813;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary proline-rich protein PO (Allele M) [Contains: Peptide P-D] (Fragment).  
DE (Fragment).  
GN PRB4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89121439; PubMed=3220251;  
RA Lyons K.M., Stein J.H., Smithies O.;

RT "Many protein products from a few loci: assignment of human salivary proline-rich proteins to specific loci." Genet. 120:255-265 (1988). [2]

RP SEQUENCE OF 165-234.

RX MEDLINE=83186122; PubMed=6841349;

RA Saichoh E., Isemura S., Sanada K.;

RT "Complete amino acid sequence of a basic proline-rich peptide, P-D, from human parotid saliva." J. Biochem. 93:495-502 (1983).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X07704; CAA30542.1; -

DR MIM; 168730; -

DR MIM; 180990; -

DR GO; GO:0005576; C:extracellular; NBS.

KW Repeat; Parotid gland; Multigene family.

FT NON TER 1 234  
FT CHAIN 165 234  
FT PEPTIDE P-D.

SQ SEQUENCE 234 AA; 23676 MW; 310AFF13A44E747F CRC64;

Query Match 84.9%; Score 45; DB 1; Length 234;  
Best Local Similarity 87.5%; Pred. No. 2.5;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGPQ 9

Db 168 GKPQGPQ 175

## RESULT 3

PRP2\_HUMAN STANDARD; PRT; 247 AA.

AC P10163; P02813;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary proline-rich protein PO precursor (Allele S).  
GN PRB4.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.

RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the  
human salivary proline-rich protein gene system."  
RL J. Biol. Chem. 260:11123-11130(1985).

RN [2]

RP SEQUENCE FROM N.A.

RA Lyons K.M., Stein J.H., Smithies O.;  
RT "Length polymorphisms in human proline-rich protein genes generated  
by intragenic unequal crossing over."  
RL Genetics 120:267-278(1988).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; K03207; AAA60188.1; -

PIR; S03176; PIHUSD.

GeneW; HGNC:9340; PRB4.

MIN; 168730; -

MIM; 180990; -

FT CHAIN 17 247 SALIVARY PROLINE-RICH PROTEIN PO.  
FT CHAIN 17 39 PROTEIN NL.  
FT CHAIN 40 177 GLYCOSYLATED PROTEIN A.  
FT CHAIN 37 57 MISSING (IN REF. 2).  
FT CONFLICT 218 218 D -> A (IN REF. 2).

SQ SEQUENCE 247 AA; 25108 MW; 6A1943E435161691 CRC64;

Query Match 84.9%; Score 45; DB 1; Length 247;  
Best Local Similarity 87.5%; Pred. No. 2.6;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGPQ 9

Db 181 GKPQGPQ 188

## RESULT 4

PRP2\_HUMAN

PRP2\_HUMAN STANDARD; PRT; 251 AA.

AC P02812;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary proline-rich protein precursor (Clone CP7) [Contains: Basic  
peptide P-F] (Fragment).  
GN PRB2.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.

RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the  
human salivary proline-rich protein gene system."  
RL J. Biol. Chem. 260:11123-11130(1985).

RN [2]

RP SEQUENCE OF 134-194.

RA Saich E., Isemura S., Sanada K.;

RT "Complete amino acid sequence of a basic proline-rich peptide, P-F,  
from human parotid saliva."  
RL J. Biochem. 93:883-888(1983).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; K03208; AAA60189.1; -

PIR; E25372; PIHUPF.

GeneW; HGNC:9338; PRB2.

MIM; 168810; -

FT CHAIN 134 194 BASIC PEPTIDE P-F.

FT REPEAT <1 9

FT REPEAT 10 71

FT REPEAT 72 133

FT REPEAT 134 195

SQ SEQUENCE 251 AA; 24641 MW; D779F590C0EBF30B CRC64;

Query Match 84.9%; Score 45; DB 1; Length 251;

Best Local Similarity 87.5%; Pred. No. 2.6;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGPQ 9

Db 199 GKPQGPQ 206

RESULT 5

PRP2\_HUMAN STANDARD; PRT; 276 AA.

AC P10162; P02813;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Salivary proline-rich protein PO (Allele K) [Contains: Peptide P-D]

GN PRB4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89121439; PubMed=3220251;  
 RA Lyons K.M., Stein J.H., Smithies O.;  
 RT "Many protein products from a few loci: assignment of human salivary  
 RT proline-rich proteins to specific loci."  
 RL Genetics 120:255-265(1988).  
 [2]  
 RN SEQUENCE OF 207-276.  
 RP MEDLINE=83186122; PubMed=6841349;  
 RX Saitho E., Isemura S., Sanada K.;  
 RA "Complete amino acid sequence of a basic proline-rich peptide, P-D,  
 RT from human parotid saliva."  
 RL J. Biochem. 93:495-502(1983).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X07715; CAA30543.1; ALT\_SEQ.  
 DR PIR; S03176; PIHUSD.  
 DR MIM; 168730; -  
 DR MIM; 180990; -  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR InterPro; IPR000637; AI-HOOK.  
 DR PRINTS; PR00929; ARHOOK.  
 DR Repeat; Parotid gland; Multigene family.  
 KW NON TER 1  
 FT CHAIN 207 276 PEPTIDE P-D.  
 FT SEQUENCE 276 AA; 27816 MW; 9F49426C979441A CRC64;  
 SQ  
 Query Match 84.9%; Score 45; DB 1; Length 276;  
 Best Local Similarity 87.5%; Pred. No. 2.9;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GRPQGPQ 9  
 DB 210 GRPQGPQ 217  
 |.|||||  
 RESULT 6  
 PRP1 HUMAN  
 ID PRP1\_HUMAN STANDARD; PRT; 331 AA.  
 AC P04280;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Salivary proline-rich protein precursor (Clones CP3, CP4 and CP5)  
 DE [Contains: Basic peptide IB-6; Peptide P-H].  
 GN PRB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85289325; PubMed=2993301;  
 RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
 RT "Differential RNA splicing and post-translational cleavages in the  
 RT human salivary proline-rich protein gene system."  
 RL J. Biol. Chem. 260:11123-11130(1985).  
 [2]  
 RN SEQUENCE OF 214-331.  
 RP MEDLINE=86243355; PubMed=3521730;  
 RX Kaufman D., Hofmann T., Bennick A., Keller P.;  
 RA "Basic proline-rich proteins from human parotid saliva: complete  
 RT covalent structures of proteins IB-1 and IB-6."  
 RL Biochemistry 25:2387-2392(1986).  
 [3]  
 RN SEQUENCE OF 276-331.  
 RP MEDLINE=84161824; PubMed=6671974;

RA Saitho E., Isemura S., Sanada K.;  
 RT "Further fractionation of basic proline-rich peptides from human  
 RT parotid saliva and complete amino acid sequence of basic proline-rich  
 RT peptide P-H."  
 RL J. Biochem. 94:1991-1997(1983).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X03204; AAA60185.1; -  
 DR EMBL; X03205; AAA60186.1; -  
 DR EMBL; X03206; AAA60187.1; -  
 DR PIR; B40750; PIHUB6.  
 DR Genew; HGNC:9337; PRB1.  
 DR MIM; 180989; -  
 DR Repeat; Parotid gland; Multigene family; Signal.  
 KW SIGNAL 1 16  
 FT CHAIN 214 331 PEPTIDE IB-6.  
 FT CHAIN 276 331 PEPTIDE P-H.  
 FT VARIANT 106 238 Missing (in clone CP-4).  
 FT VARIANT 106 258 /FTid=VAR\_005561.  
 FT CONFLICT 276 276 Missing (in clone CP-5).  
 FT SEQUENCE 331 AA; 32596 MW; 3F481FF8EBA39751 CRC64;  
 SQ  
 Query Match 84.9%; Score 45; DB 1; Length 331;  
 Best Local Similarity 87.5%; Pred. No. 3.5;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GRPQGPQ 9  
 DB 177 GRPQGPQ 184  
 |.|||||  
 RESULT 7  
 CES4 HUMAN  
 ID CES4\_HUMAN STANDARD; PRT; 124 AA.  
 AC Q9BXQ7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cat eye syndrome critical region protein 4 (Fragment).  
 GN CECR4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21275466; PubMed=11381032;  
 RA Footz T.K., Brinkman-Mills P., Banting G.S., Maier S.A., Riazi M.A.,  
 RA Bridgland L.J., Hu S., Birren B., Minoshima S., Shimizu N., Pan H.,  
 RA Nguyen T., Fang F., Fu Y., Ray L., Wu H., Shaul S., Phan S., Yao Z.,  
 RA Chen F., Huan A., Hu P., Wang Q., Loh P., Qi S., Roe B.A.,  
 RA McDermid H.E.;  
 RT "Analysis of the cat eye syndrome critical region in humans and the  
 RT region of conserved synteny in mice: a search for candidate genes at  
 RT or near the human chromosome 22 pericentromere."  
 RL Genome Res. 11:1053-1070(2001).  
 CC -!- TISSUE SPECIFICITY: Adult heart and skeletal muscle. Widely  
 CC expressed in fetal tissues.  
 CC -!- MISCELLANEOUS: Candidate gene for the Cat Eye Syndrome (CES), a  
 CC developmental disorder associated with the duplication of a 2 Mb  
 CC region of 22q11.2. Duplication usually takes in the form of a  
 CC supernumerary bisatellited isodicentric chromosome, resulting in  
 CC four copies of the region (represents an inv dup(22)(q11)). CES is  
 CC characterized clinically by the combination of coloboma of the



iris and anal atresia with fistula, downslanting palpebral fissures, preauricular tags and/or pits, frequent occurrence of heart and renal malformations, and normal or near-normal mental development.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF307448; AAK30048.1; ..  
Genew; HGNC:1842; CECR4.  
NON\_TER 1  
SEQUENCE 124 AA; 12838 MW; 938E00386308EC7A CRC64;

Query Match 79.2%; Score 42; DB 1; Length 124;  
Best Local Similarity 87.5%; Pred. No. 3.8;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGQPPP 8  
|||  
58 RGRPGQAPP 65

Db

RESULT 8  
PRP3 RAT  
ID PRP3 RAT STANDARD; PRT; 206 AA.  
AC P04474;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Acidic proline-rich protein PR233 precursor.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
[1]  
SEQUENCE FROM N.A.  
RP TISSUE=Parotid gland;  
RX MEDLINE=84289443; PubMed=6547951;  
RA Ziemer M.A., Swain W.F., Rutter W.J., Clements S., Ann D.K.,  
RA Carlson D.M.;  
RT "Nucleotide sequence analysis of a proline-rich protein cDNA and  
RL peptide homologues of rat and human proline-rich proteins.";  
RL J. Biol. Chem. 259:10475-10480(1984).  
[2]  
SEQUENCE OF 1-23 FROM N.A.  
RP TISSUE=Parotid gland;  
RX MEDLINE=86033799; PubMed=3840480;  
RA Clements S., Mehansho H., Carlson D.M.;  
RT "Novel multigene families encoding highly repetitive peptide  
RT sequences. Sequence analyses of rat and mouse proline-rich protein  
RT cDNAs.";  
RL J. Biol. Chem. 260:13471-13477(1985).  
CC  
CC -!- FUNCTION: May protect teeth by binding to tannins.  
-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; K02247; AAA41949.1; ..  
DR EMBL; M11898; AAA41958.1; ..  
DR PIR; A03296; PIRT3.  
KW Repeat; Parotid gland; Multigene family; Signal.  
FT SIGNAL 1 13  
POTENTIAL.  
FT CHAIN 14 206  
ACIDIC PROLINE-RICH PROTEIN PRP33.

FT DOMAIN 80 189 6 X 18 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 80 97 1.  
FT REPEAT 98 115 2.  
FT REPEAT 116 133 3.  
FT REPEAT 134 152 4.  
FT REPEAT 153 170 5.  
FT REPEAT 171 189 6.  
SQ SEQUENCE 206 AA; D037582CED05CE6B CRC64;

Query Match 79.2%; Score 42; DB 1; Length 206;  
Best Local Similarity 77.8%; Pred. No. 6.3;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGQPPP 9  
|||  
119 QGGPGQAPP 127

Db

RESULT 9  
PAT1 YEAST  
ID PAT1 YEAST STANDARD; PRT; 797 AA.  
AC P25444;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Topoisomerase II-associated protein PAT1.  
GN PAT1 OR YCR077C OR YCR77C.  
OS Saccharomyces cerevisiae (Baker's Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
[1]  
SEQUENCE FROM N.A.  
RP Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,  
RA Sanz E.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
REVIEWS, AND CHARACTERIZATION.  
RX MEDLINE=95373282; PubMed=7645349;  
RA Rodriguez-Cousino N., Lill R., Neupert W., Court D.A.;  
RT "Identification and initial characterization of the cytosolic protein  
RT Ycr77p.";  
RL Yeast 11:581-585 (1995).  
RN [3]  
FUNCTION.  
RP MEDLINE=97128274; PubMed=8972867;  
RX Wang X., Watt P.M., Louis E.J., Borts R.H., Hickson I.D.;  
RT "Pat1: a topoisomerase II-associated protein required for faithful  
RT chromosome transmission in Saccharomyces cerevisiae.";  
RL Nucleic Acids Res. 24:4791-4797(1996).  
CC -!- FUNCTION: Necessary for accurate chromosome transmission during  
CC cell division. Interacts with topoisomerase TOP2.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: TO S.POMBE SPBC19G7.10C.  
-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; X59720; CAA42266.1; ..  
DR PIR; S53590; S53590.  
DR GerMOnline; 138977; ..  
DR SGD; S0000673; PAT1.  
DR GO; GO:0000932; C:cytoplasmic mRNA processing body; IDA.  
DR GO; GO:0005843; C:cytosolic small ribosomal subunit (sensu Eu. . .; IDA.  
DR GO; GO:0007059; P:chromosome segregation; IMP.  
DR GO; GO:0006446; P:regulation of translational initiation; IGI.  
DOMAIN 195 200  
POLY-PRO.  
FT SEQUENCE 797 AA; 88487 MW; 2A54F39AE3E75ECE CRC64;



Query Match 79.2%; Score 42; DB 1; Length 797;  
 Best Local Similarity 77.8%; Pred. No. 24;  
 Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 1 RGRPGPPQ 9  
 : |||||  
 DB 168 QGLFQGPQ 176

## RESULT 10

ID SYB LOLPE STANDARD; PRT; 125 AA.  
 AC P47194;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Synaptobrevin.  
 OS Loligo pealeii (longfin squid).  
 OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;  
 OC Decapodiformes; Loliginidae; Loligo.  
 OX NCBI\_TaxID=6621;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94280860; PubMed=8011337;  
 RA Hunt J.M., Bommer K., Charlton M.P., Kistner A., Habermann E.,  
 RA Augustine G.J., Betz H.;  
 RT "A post-docking role for synaptobrevin in synaptic vesicle fusion.";  
 RL Neuron 12:1269-1279(1994).  
 CC -!- FUNCTION: Intrinsic membrane protein of small synaptic  
 CC vesicles.  
 CC -!- SUBCELLULAR LOCATION: Type IV membrane protein. Neuronal synaptic  
 CC vesicles.  
 CC -!- SIMILARITY: Belongs to the synaptobrevin family.  
 CC -!- SIMILARITY: Contains 1 v-SNARE coiled-coil homology domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X74748; CAA52766.1; --  
 CC PIR; S40153;  
 CC PDB; 1L4A; 31-JUL-02.  
 CC InterPro: IPR001388; Synaptobrevin.  
 CC Pfam; PF00957; Synaptobrevin; 1.  
 CC PRINTS; PR00219; SYNAPTOBREVN.  
 CC ProDom; PD001229; Synaptobrevin; 1.  
 CC PROSITE; PS00417; SYNAPTOBREVN; 1.  
 CC PROSITE; PS50892; V SNARE; 1.  
 CC Synapse; Synaptosome; Transmembrane; Coiled coil; 3D-structure.  
 CC DOMAIN 1 103 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 104 123 ANCHOR FOR TYPE IV MEMBRANE PROTEIN  
 CC (POTENTIAL).  
 CC FT DOMAIN 124 125 VESICULAR (POTENTIAL).  
 CC FT DOMAIN 40 100 V-SNARE COILED-COIL HOMOMLOGY.  
 CC SEQUENCE 125 AA; 13316 MW; 6FE9E5604ECASD86 CRC64;

Query Match 77.4%; Score 41; DB 1; Length 125;  
 Best Local Similarity 87.5%; Pred. No. 5;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPGQPPQ 9  
 : |||||  
 DB 26 GPPQGPQ 33

## RESULT 11

ID SCAP\_CRIGR STANDARD; PRT; 1276 AA.

P97260;  
 AC 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Sterol regulatory element binding protein cleavage-activating protein  
 DE (SREBP cleavage-activating protein) (SCAP).  
 GN SCAP.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetulus.  
 OX NCBI\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS OF ASP-443.  
 RC TISSUE=Ovary;  
 RX MEDLINE=97053781; PubMed=8898195;  
 RA Hua X., Nonturff A., Goldstein J.L., Brown M.S.;  
 RT "Sterol resistance in CHO cells traced to point mutation in SREBP  
 RT cleavage-activating protein.";  
 RL Cell 87:415-426(1996).  
 CC [2]  
 RN [2]  
 RP POST-TRANSLATIONAL MODIFICATIONS, AND MUTAGENESIS OF TYR-298.  
 RC TISSUE=Ovary;  
 RX MEDLINE=99432216; PubMed=10500160;  
 RA Nonturff A., DeBose-Boyd R.A., Scheek S., Goldstein J.L., Brown M.S.;  
 RT "Sterols regulate cycling of SREBP cleavage-activating protein (SCAP)  
 RT between endoplasmic reticulum and Golgi.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11235-11240(1999).  
 CC [3]  
 RN [3]  
 RP TOPOLOGY, AND CARBOHYDRATE-LINKAGE SITES ASN-263; ASN-590 AND ASN-641.  
 RX MEDLINE=98307971; PubMed=9642295;  
 RA Nonturff A., Brown M.S., Goldstein J.L.;  
 RT "Topology of SREBP cleavage-activating protein, a polytopic membrane  
 RT protein with a sterol-sensing domain.";  
 RL J. Biol. Chem. 273:17243-17250(1998).  
 CC [4]  
 RN [4]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=20449077; PubMed=10896675;  
 RA Yang T., Goldstein J.L., Brown M.S.;  
 RT "Overexpression of membrane domain of SCAP prevents sterols from  
 RT inhibiting SCAP.SREBP exit from endoplasmic reticulum.";  
 RL J. Biol. Chem. 275:29881-29886(2000).  
 CC -!- FUNCTION: Sterol sensor. Necessary for the proteolytic activation  
 CC of SREBPs by site-1 protease in the Golgi.  
 CC -!- SUBUNIT: In a tight complex with SREBs.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Moves from the  
 CC endoplasmic reticulum to the Golgi in the absence of sterols.  
 CC -!- SIMILARITY: Contains 1 sterol sensor (SSD) domain.  
 CC -!- SIMILARITY: Contains 7 WD repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U67060; AAB19103.1; --  
 CC PIR; T18526; T18526.  
 CC GO; GO:0005783; C:endoplasmic reticulum; ISS.  
 CC GO; GO:0005794; C:Golgi apparatus; ISS.  
 CC GO; GO:0016021; C:integral to membrane; ISS.  
 CC GO; GO:0003754; F:chaperone activity; NAS.  
 CC GO; GO:0045541; F:negative regulation of cholesterol biosynth. . . ; ISS.  
 CC GO; GO:0045716; F:positive regulation of low-density lipoprot. . . ; ISS.  
 CC GO; GO:0006994; P:sterol depletion response, SREBP target gen. . . ; ISS.  
 CC InterPro; IPR000731; SSD 5TM.  
 CC InterPro; IPR001680; WD40.  
 CC Pfam; PF00400; WD40; 4.  
 CC SMART; SM00320; WD40; 6.  
 CC PROSITE; PS00156; SSD; 1.  
 CC PROSITE; PS00678; WD\_REPEATS\_1; 1.

```

DR PROSITE; PS50082; WD REPEATS 2; 1.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
KW Lipid metabolism; Cholesterol metabolism; Transmembrane;
KW Endoplasmic reticulum; Golgi stack; Repeat; WD repeat; Glycoprotein.
FT DOMAIN 1 18
FT TRANSMEM 19 39
FT DOMAIN 40 279
FT TRANSMEM 280 300
FT TRANSMEM 301 312
FT TRANSMEM 313 323
FT TRANSMEM 324 334
FT TRANSMEM 335 344
FT TRANSMEM 345 365
FT TRANSMEM 366 401
FT TRANSMEM 402 422
FT TRANSMEM 423 444
FT TRANSMEM 445 518
FT TRANSMEM 519 539
FT TRANSMEM 540 708
FT TRANSMEM 709 729
FT TRANSMEM 730 1276
FT TRANSMEM 1277 1296
FT TRANSMEM 1297 1317
FT TRANSMEM 1318 1338
FT TRANSMEM 1339 1359
FT TRANSMEM 1360 1380
FT TRANSMEM 1381 1401
FT TRANSMEM 1402 1422
FT TRANSMEM 1423 1443
FT TRANSMEM 1444 1464
FT TRANSMEM 1465 1485
FT TRANSMEM 1486 1506
FT TRANSMEM 1507 1527
FT TRANSMEM 1528 1548
FT TRANSMEM 1549 1569
FT TRANSMEM 1570 1590
FT TRANSMEM 1591 1610
FT TRANSMEM 1611 1631
FT TRANSMEM 1632 1652
FT TRANSMEM 1653 1673
FT TRANSMEM 1674 1694
FT TRANSMEM 1695 1715
FT TRANSMEM 1716 1736
FT TRANSMEM 1737 1757
FT TRANSMEM 1758 1778
FT TRANSMEM 1779 1799
FT TRANSMEM 1800 1820
FT TRANSMEM 1821 1841
FT TRANSMEM 1842 1862
FT TRANSMEM 1863 1883
FT TRANSMEM 1884 1904
FT TRANSMEM 1905 1925
FT TRANSMEM 1926 1946
FT TRANSMEM 1947 1967
FT TRANSMEM 1968 1988
FT TRANSMEM 1989 2009
FT TRANSMEM 2010 2030
FT TRANSMEM 2031 2051
FT TRANSMEM 2052 2072
FT TRANSMEM 2073 2093
FT TRANSMEM 2094 2114
FT TRANSMEM 2115 2135
FT TRANSMEM 2136 2156
FT TRANSMEM 2157 2177
FT TRANSMEM 2178 2198
FT TRANSMEM 2199 2219
FT TRANSMEM 2220 2240
FT TRANSMEM 2241 2261
FT TRANSMEM 2262 2282
FT TRANSMEM 2283 2303
FT TRANSMEM 2304 2324
FT TRANSMEM 2325 2345
FT TRANSMEM 2346 2366
FT TRANSMEM 2367 2387
FT TRANSMEM 2388 2408
FT TRANSMEM 2409 2429
FT TRANSMEM 2430 2450
FT TRANSMEM 2451 2471
FT TRANSMEM 2472 2492
FT TRANSMEM 2493 2513
FT TRANSMEM 2514 2534
FT TRANSMEM 2535 2555
FT TRANSMEM 2556 2576
FT TRANSMEM 2577 2597
FT TRANSMEM 2598 2618
FT TRANSMEM 2619 2639
FT TRANSMEM 2640 2660
FT TRANSMEM 2661 2681
FT TRANSMEM 2682 2702
FT TRANSMEM 2703 2723
FT TRANSMEM 2724 2744
FT TRANSMEM 2745 2765
FT TRANSMEM 2766 2786
FT TRANSMEM 2787 2807
FT TRANSMEM 2808 2828
FT TRANSMEM 2829 2849
FT TRANSMEM 2850 2870
FT TRANSMEM 2871 2891
FT TRANSMEM 2892 2912
FT TRANSMEM 2913 2933
FT TRANSMEM 2934 2954
FT TRANSMEM 2955 2975
FT TRANSMEM 2976 2996
FT TRANSMEM 2997 3017
FT TRANSMEM 3018 3038
FT TRANSMEM 3039 3059
FT TRANSMEM 3060 3080
FT TRANSMEM 3081 3101
FT TRANSMEM 3102 3122
FT TRANSMEM 3123 3143
FT TRANSMEM 3144 3164
FT TRANSMEM 3165 3185
FT TRANSMEM 3186 3206
FT TRANSMEM 3207 3227
FT TRANSMEM 3228 3248
FT TRANSMEM 3249 3269
FT TRANSMEM 3270 3290
FT TRANSMEM 3291 3311
FT TRANSMEM 3312 3332
FT TRANSMEM 3333 3353
FT TRANSMEM 3354 3374
FT TRANSMEM 3375 3395
FT TRANSMEM 3396 3416
FT TRANSMEM 3417 3437
FT TRANSMEM 3438 3458
FT TRANSMEM 3459 3479
FT TRANSMEM 3480 3500
FT TRANSMEM 3501 3521
FT TRANSMEM 3522 3542
FT TRANSMEM 3543 3563
FT TRANSMEM 3564 3584
FT TRANSMEM 3585 3605
FT TRANSMEM 3606 3626
FT TRANSMEM 3627 3647
FT TRANSMEM 3648 3668
FT TRANSMEM 3669 3689
FT TRANSMEM 3690 3710
FT TRANSMEM 3711 3731
FT TRANSMEM 3732 3752
FT TRANSMEM 3753 3773
FT TRANSMEM 3774 3794
FT TRANSMEM 3795 3815
FT TRANSMEM 3816 3836
FT TRANSMEM 3837 3857
FT TRANSMEM 3858 3878
FT TRANSMEM 3879 3899
FT TRANSMEM 3900 3920
FT TRANSMEM 3921 3941
FT TRANSMEM 3942 3962
FT TRANSMEM 3963 3983
FT TRANSMEM 3984 4004
FT TRANSMEM 4005 4025
FT TRANSMEM 4026 4046
FT TRANSMEM 4047 4067
FT TRANSMEM 4068 4088
FT TRANSMEM 4089 4109
FT TRANSMEM 4110 4130
FT TRANSMEM 4131 4151
FT TRANSMEM 4152 4172
FT TRANSMEM 4173 4193
FT TRANSMEM 4194 4214
FT TRANSMEM 4215 4235
FT TRANSMEM 4236 4256
FT TRANSMEM 4257 4277
FT TRANSMEM 4278 4298
FT TRANSMEM 4299 4319
FT TRANSMEM 4320 4340
FT TRANSMEM 4341 4361
FT TRANSMEM 4362 4382
FT TRANSMEM 4383 4403
FT TRANSMEM 4404 4424
FT TRANSMEM 4425 4445
FT TRANSMEM 4446 4466
FT TRANSMEM 4467 4487
FT TRANSMEM 4488 4508
FT TRANSMEM 4509 4529
FT TRANSMEM 4530 4550
FT TRANSMEM 4551 4571
FT TRANSMEM 4572 4592
FT TRANSMEM 4593 4613
FT TRANSMEM 4614 4634
FT TRANSMEM 4635 4655
FT TRANSMEM 4656 4676
FT TRANSMEM 4677 4697
FT TRANSMEM 4698 4718
FT TRANSMEM 4719 4739
FT TRANSMEM 4740 4760
FT TRANSMEM 4761 4781
FT TRANSMEM 4782 4802
FT TRANSMEM 4803 4823
FT TRANSMEM 4824 4844
FT TRANSMEM 4845 4865
FT TRANSMEM 4866 4886
FT TRANSMEM 4887 4907
FT TRANSMEM 4908 4928
FT TRANSMEM 4929 4949
FT TRANSMEM 4950 4970
FT TRANSMEM 4971 4991
FT TRANSMEM 4992 5012
FT TRANSMEM 5013 5033
FT TRANSMEM 5034 5054
FT TRANSMEM 5055 5075
FT TRANSMEM 5076 5096
FT TRANSMEM 5097 5117
FT TRANSMEM 5118 5138
FT TRANSMEM 5139 5159
FT TRANSMEM 5160 5180
FT TRANSMEM 5181 5201
FT TRANSMEM 5202 5222
FT TRANSMEM 5223 5243
FT TRANSMEM 5244 5264
FT TRANSMEM 5265 5285
FT TRANSMEM 5286 5306
FT TRANSMEM 5307 5327
FT TRANSMEM 5328 5348
FT TRANSMEM 5349 5369
FT TRANSMEM 5370 5390
FT TRANSMEM 5391 5411
FT TRANSMEM 5412 5432
FT TRANSMEM 5433 5453
FT TRANSMEM 5454 5474
FT TRANSMEM 5475 5495
FT TRANSMEM 5496 5516
FT TRANSMEM 5517 5537
FT TRANSMEM 5538 5558
FT TRANSMEM 5559 5579
FT TRANSMEM 5580 5600
FT TRANSMEM 5601 5621
FT TRANSMEM 5622 5642
FT TRANSMEM 5643 5663
FT TRANSMEM 5664 5684
FT TRANSMEM 5685 5705
FT TRANSMEM 5706 5726
FT TRANSMEM 5727 5747
FT TRANSMEM 5748 5768
FT TRANSMEM 5769 5789
FT TRANSMEM 5790 5810
FT TRANSMEM 5811 5831
FT TRANSMEM 5832 5852
FT TRANSMEM 5853 5873
FT TRANSMEM 5874 5894
FT TRANSMEM 5895 5915
FT TRANSMEM 5916 5936
FT TRANSMEM 5937 5957
FT TRANSMEM 5958 5978
FT TRANSMEM 5979 5999
FT TRANSMEM 6000 6021
FT TRANSMEM 6022 6042
FT TRANSMEM 6043 6063
FT TRANSMEM 6064 6084
FT TRANSMEM 6085 6105
FT TRANSMEM 6106 6126
FT TRANSMEM 6127 6147
FT TRANSMEM 6148 6168
FT TRANSMEM 6169 6189
FT TRANSMEM 6190 6210
FT TRANSMEM 6211 6231
FT TRANSMEM 6232 6252
FT TRANSMEM 6253 6273
FT TRANSMEM 6274 6294
FT TRANSMEM 6295 6315
FT TRANSMEM 6316 6336
FT TRANSMEM 6337 6357
FT TRANSMEM 6358 6378
FT TRANSMEM 6379 6399
FT TRANSMEM 6400 6421
FT TRANSMEM 6422 6442
FT TRANSMEM 6443 6463
FT TRANSMEM 6464 6484
FT TRANSMEM 6485 6505
FT TRANSMEM 6506 6526
FT TRANSMEM 6527 6547
FT TRANSMEM 6548 6568
FT TRANSMEM 6569 6589
FT TRANSMEM 6590 6610
FT TRANSMEM 6611 6631
FT TRANSMEM 6632 6652
FT TRANSMEM 6653 6673
FT TRANSMEM 6674 6694
FT TRANSMEM 6695 6715
FT TRANSMEM 6716 6736
FT TRANSMEM 6737 6757
FT TRANSMEM 6758 6778
FT TRANSMEM 6779 6799
FT TRANSMEM 6800 6821
FT TRANSMEM 6822 6842
FT TRANSMEM 6843 6863
FT TRANSMEM 6864 6884
FT TRANSMEM 6885 6905
FT TRANSMEM 6906 6926
FT TRANSMEM 6927 6947
FT TRANSMEM 6948 6968
FT TRANSMEM 6969 6989
FT TRANSMEM 6990 7010
FT TRANSMEM 7011 7031
FT TRANSMEM 7032 7052
FT TRANSMEM 7053 7073
FT TRANSMEM 7074 7094
FT TRANSMEM 7095 7115
FT TRANSMEM 7116 7136
FT TRANSMEM 7137 7157
FT TRANSMEM 7158 7178
FT TRANSMEM 7179 7199
FT TRANSMEM 7200 7221
FT TRANSMEM 7222 7242
FT TRANSMEM 7243 7263
FT TRANSMEM 7264 7284
FT TRANSMEM 7285 7305
FT TRANSMEM 7306 7326
FT TRANSMEM 7327 7347
FT TRANSMEM 7348 7368
FT TRANSMEM 7369 7389
FT TRANSMEM 7390 7410
FT TRANSMEM 7411 7431
FT TRANSMEM 7432 7452
FT TRANSMEM 7453 7473
FT TRANSMEM 7474 7494
FT TRANSMEM 7495 7515
FT TRANSMEM 7516 7536
FT TRANSMEM 7537 7557
FT TRANSMEM 7558 7578
FT TRANSMEM 7579 7599
FT TRANSMEM 7600 7621
FT TRANSMEM 7622 7642
FT TRANSMEM 7643 7663
FT TRANSMEM 7664 7684
FT TRANSMEM 7685 7705
FT TRANSMEM 7706 7726
FT TRANSMEM 7727 7747
FT TRANSMEM 7748 7768
FT TRANSMEM 7769 7789
FT TRANSMEM 7790 7810
FT TRANSMEM 7811 7831
FT TRANSMEM 7832 7852
FT TRANSMEM 7853 7873
FT TRANSMEM 7874 7894
FT TRANSMEM 7895 7915
FT TRANSMEM 7916 7936
FT TRANSMEM 7937 7957
FT TRANSMEM 7958 7978
FT TRANSMEM 7979 7999
FT TRANSMEM 8000 8021
FT TRANSMEM 8022 8042
FT TRANSMEM 8043 8063
FT TRANSMEM 8064 8084
FT TRANSMEM 8085 8105
FT TRANSMEM 8106 8126
FT TRANSMEM 8127 8147
FT TRANSMEM 8148 8168
FT TRANSMEM 8169 8189
FT TRANSMEM 8190 8210
FT TRANSMEM 8211 8231
FT TRANSMEM 8232 8252
FT TRANSMEM 8253 8273
FT TRANSMEM 8274 8294
FT TRANSMEM 8295 8315
FT TRANSMEM 8316 8336
FT TRANSMEM 8337 8357
FT TRANSMEM 8358 8378
FT TRANSMEM 8379 8399
FT TRANSMEM 8400 8421
FT TRANSMEM 8422 8442
FT TRANSMEM 8443 8463
FT TRANSMEM 8464 8484
FT TRANSMEM 8485 8505
FT TRANSMEM 8506 8526
FT TRANSMEM 8527 8547
FT TRANSMEM 8548 8568
FT TRANSMEM 8569 8589
FT TRANSMEM 8590 8610
FT TRANSMEM 8611 8631
FT TRANSMEM 8632 8652
FT TRANSMEM 8653 8673
FT TRANSMEM 8674 8694
FT TRANSMEM 8695 8715
FT TRANSMEM 8716 8736
FT TRANSMEM 8737 8757
FT TRANSMEM 8758 8778
FT TRANSMEM 8779 8799
FT TRANSMEM 8800 8821
FT TRANSMEM 8822 8842
FT TRANSMEM 8843 8863
FT TRANSMEM 8864 8884
FT TRANSMEM 8885 8905
FT TRANSMEM 8906 8926
FT TRANSMEM 8927 8947
FT TRANSMEM 8948 8968
FT TRANSMEM 8969 8989
FT TRANSMEM 8990 9010
FT TRANSMEM 9011 9031
FT TRANSMEM 9032 9052
FT TRANSMEM 9053 9073
FT TRANSMEM 9074 9094
FT TRANSMEM 9095 9115
FT TRANSMEM 9116 9136
FT TRANSMEM 9137 9157
FT TRANSMEM 9158 9178
FT TRANSMEM 9179 9199
FT TRANSMEM 9200 9221
FT TRANSMEM 9222 9242
FT TRANSMEM 9243 9263
FT TRANSMEM 9264 9284
FT TRANSMEM 9285 9305
FT TRANSMEM 9306 9326
FT TRANSMEM 9327 9347
FT TRANSMEM 9348 9368
FT TRANSMEM 9369 9389
FT TRANSMEM 9390 9410
FT TRANSMEM 9411 9431
FT TRANSMEM 9432 9452
FT TRANSMEM 9453 9473
FT TRANSMEM 9474 9494
FT TRANSMEM 9495 9515
FT TRANSMEM 9516 9536
FT TRANSMEM 9537 9557
FT TRANSMEM 9558 9578
FT TRANSMEM 9579 9599
FT TRANSMEM 9600 9621
FT TRANSMEM 9622 9642
FT TRANSMEM 9643 9663
FT TRANSMEM 9664 9684
FT TRANSMEM 9685 9705
FT TRANSMEM 9706 9726
FT TRANSMEM 9727 9747
FT TRANSMEM 9748 9768
FT TRANSMEM 9769 9789
FT TRANSMEM 9790 9810
FT TRANSMEM 9811 9831
FT TRANSMEM 9832 9852
FT TRANSMEM 9853 9873
FT TRANSMEM 9874 9894
FT TRANSMEM 9895 9915
FT TRANSMEM 9916 9936
FT TRANSMEM 9937 9957
FT TRANSMEM 9958 9978
FT TRANSMEM 9979 9999
FT TRANSMEM 10000 10021
FT TRANSMEM 10022 10042
FT TRANSMEM 10043 10063
FT TRANSMEM 10064 10084
FT TRANSMEM 10085 10105
FT TRANSMEM 10106 10126
FT TRANSMEM 10127 10147
FT TRANSMEM 10148 10168
FT TRANSMEM 10169 10189
FT TRANSMEM 10190 10210
FT TRANSMEM 10211 10231
FT TRANSMEM 10232 10252
FT TRANSMEM 10253 10273
FT TRANSMEM 10274 10294
FT TRANSMEM 10295 10315
FT TRANSMEM 10316 10336
FT TRANSMEM 10337 10357
FT TRANSMEM 10358 10378
FT TRANSMEM 10379 10399
FT TRANSMEM 10400 10421
FT TRANSMEM 10422 10442
FT TRANSMEM 10443 10463
FT TRANSMEM 10464 10484
FT TRANSMEM 10485 10505
FT TRANSMEM 10506 10526
FT TRANSMEM 10527 10547
FT TRANSMEM 10548 10568
FT TRANSMEM 10569 10589
FT TRANSMEM 10590 10610
FT TRANSMEM 10611 10631
FT TRANSMEM 10632 10652
FT TRANSMEM 10653 10673
FT TRANSMEM 10674 10694
FT TRANSMEM 10695 10715
FT TRANSMEM 10716 10736
FT TRANSMEM 10737 10757
FT TRANSMEM 10758 10778
FT TRANSMEM 10779 10799
FT TRANSMEM 10800 10821
FT TRANSMEM 10822 10842
FT TRANSMEM 10843 10863
FT TRANSMEM 10864 10884
FT TRANSMEM 10885 10905
FT TRANSMEM 10906 10926
FT TRANSMEM 10927 10947
FT TRANSMEM 10948 10968
FT TRANSMEM 10969 10989
FT TRANSMEM 10990 11010
FT TRANSMEM 11011 11031
FT TRANSMEM 11032 11052
FT TRANSMEM 11053 11073
FT TRANSMEM 11074 11094
FT TRANSMEM 11095 11115
FT TRANSMEM 11116 11136
FT TRANSMEM 11137 11157
FT TRANSMEM 11158 11178
FT TRANSMEM 11179 11199
FT TRANSMEM 11200 11221
FT TRANSMEM 11222 11242
FT TRANSMEM 11243 11263
FT TRANSMEM 11264 11284
FT TRANSMEM 11285 11305
FT TRANSMEM 11306 11326
FT TRANSMEM 11327 11347
FT TRANSMEM 11348 11368
FT TRANSMEM 11369 11389
FT TRANSMEM 11390 11410
FT TRANSMEM 11411 11431
FT TRANSMEM 11432 11452
FT TRANSMEM 11453 11473
FT TRANSMEM 11474 11494
FT TRANSMEM 11495 11515
FT TRANSMEM 11516 11536
FT TRANSMEM 11537 11557
FT TRANSMEM 11558 11578
FT TRANSMEM 11579 11599
FT TRANSMEM 11600 11621
FT TRANSMEM 11622 11642
FT TRANSMEM 11643 11663
FT TRANSMEM 11664 11684
FT TRANSMEM 11685 11705
FT TRANSMEM 11706 11726
FT TRANSMEM 11727 11747
FT TRANSMEM 11748 11768
FT TRANSMEM 11769 11789
FT TRANSMEM 11790 11810
FT TRANSMEM 11811 11831
FT TRANSMEM 11832 11852
FT TRANSMEM 11853 11873
FT TRANSMEM 11874 11894
FT TRANSMEM 11895 11915
FT TRANSMEM 11916 11936
FT TRANSMEM 11937 11957
FT TRANSMEM 11958 11978
FT TRANSMEM 11979 11999
FT TRANSMEM 12000 12021
FT TRANSMEM 12022 12042
FT TRANSMEM 12043 12063
FT TRANSMEM 12064 12084
FT TRANSMEM 12085 12105
FT TRANSMEM 12106 12126
FT TRANSMEM 12127 12147
FT TRANSMEM 12148 12168
FT TRANSMEM 12169 12189
FT TRANSMEM 12190 12210
FT TRANSMEM 12211 12231
FT TRANSMEM 12232 12252
FT TRANSMEM 12253 12273
FT TRANSMEM 12274 12294
FT TRANSMEM 12295 12315
FT TRANSMEM 12316 12336
FT TRANSMEM 12337 12357
FT TRANSMEM 12358 12378
FT TRANSMEM 12379 12399
FT TRANSMEM 12400 12421
FT TRANSMEM 12422 12442
FT TRANSMEM 12443 12463
FT TRANSMEM 12464 12484
FT TRANSMEM 12485 12505
FT TRANSMEM 12506 12526
FT TRANSMEM 12527 12547
FT TRANSMEM 12548 12568
FT TRANSMEM 12569 12589
FT TRANSMEM 12590 12610
FT TRANSMEM 12611 12631
FT TRANSMEM 12632 12652
FT TRANSMEM 12653 12673
FT TRANSMEM 12674 12694
FT TRANSMEM 12695 12715
FT TRANSMEM 12716 12736
FT TRANSMEM 12737 12757
FT TRANSMEM 12758 12778
FT TRANSMEM 12779 12799
FT TRANSMEM 12800 12821
FT TRANSMEM 12822 12842
FT TRANSMEM 12843 12863
FT TRANSMEM 12864 12884
FT TRANSMEM 12885 12905
FT TRANSMEM 12906 12926
FT TRANSMEM 12927 12947
FT TRANSMEM 12948 12968
FT TRANSMEM 12969 12989
FT TRANSMEM 12990 13010
FT TRANSMEM 13011 13031
FT TRANSMEM 13032 13052
FT TRANSMEM 13053 13073
FT TRANSMEM 13074 13094
FT TRANSMEM 13095 13115
FT TRANSMEM 13116 13136
FT TRANSMEM 13137 13157
FT TRANSMEM 13158 13178
FT TRANSMEM 13179 13199
FT TRANSMEM 13200 13221
FT TRANSMEM 13222 13242
FT TRANSMEM 13243 13263
FT TRANSMEM 13264 13284
FT TRANSMEM 13285 13305
FT TRANSMEM 13306 13326
FT TRANSMEM 13327 13347
FT TRANSMEM 13348 13368
FT TRANSMEM 13369 13389
FT TRANSMEM 13390 13410
FT TRANSMEM 13411 13431
FT TRANSMEM 13432 13452
FT TRANSMEM 13453 13473
FT TRANSMEM 13474 13494
FT TRANSMEM 13495 13515
FT TRANSMEM 13516 13536
FT TRANSMEM 13537 13557
FT TRANSMEM 13558 13578
FT TRANSMEM 13579 13599
FT TRANSMEM 13600 13621
FT TRANSMEM 13622 13642
FT TRANSMEM 13643 13663
FT TRANSMEM 13664 13684
FT TRANSMEM 13685 13705
FT TRANSMEM 13706 13726
FT TRANSMEM 13727 13747
FT TRANSMEM 13748 13768
FT TRANSMEM 13769 13789
FT TRANSMEM 13790 13810
FT TRANSMEM 13811 13831
FT TRANSMEM 13832 13852
FT TRANSMEM 13853 13873
FT TRANSMEM 13874 13894
FT TRANSMEM 13895 13915
FT TRANSMEM 13916 13936
FT TRANSMEM 13937 13957
FT TRANSMEM 13958 13978
FT TRANSMEM 13979 13999
FT TRANSMEM 14000 14021
FT TRANSMEM 14022 14042
FT TRANSMEM 14043 14063
FT TRANSMEM 14064 14084
FT TRANSMEM 14085 14105
FT TRANSMEM 14106 14126
FT TRANSMEM 14127 14147
FT TRANSMEM 14148 14168
FT TRANSMEM 14169 14189
FT TRANSMEM 14190 14210
FT TRANSMEM 14211 14231
FT TRANSMEM 14232 14252
FT TRANSMEM 14253 14273
FT TRANSMEM 14274 14294
FT TRANSMEM 14295 14315
FT TRANSMEM 14316 14336
FT TRANSMEM 14337 14357
FT TRANSMEM 14358 14378
FT TRANSMEM 14379 14399
FT TRANSMEM 14400 14421
FT TRANSMEM 14422 14442
FT TRANSMEM 14443 14463
FT TRANSMEM 14464 14484
FT TRANSMEM 14485 14505
FT TRANSMEM 14506 14526
FT TRANSMEM 14527 14547
FT TRANSMEM 14548 14568
FT TRANSMEM 14569 14589
FT TRANSMEM 14590 14610
FT TRANSMEM 14611 14631
FT TRANSMEM 14632 14652
FT TRANSMEM 14653 14673
FT TRANSMEM 14674 14694
FT TRANSMEM 14695 14715
FT TRANSMEM 14716 14736
FT TRANSMEM 14737 14757
FT TRANSMEM 14758 14778
FT TRANSMEM 14779 14799
FT TRANSMEM 14800 14821
FT TRANSMEM 14822 14842
FT TRANSMEM 14843 14863
FT TRANSMEM 14864 14884
FT TRANSMEM 14885 14905
FT TRANSMEM 14906 14926
FT TRANSMEM 14927 14947
FT TRANSMEM 14948 14968
FT TRANSMEM 14969 14989
FT TRANSMEM 14990 15010
FT TRANSMEM 15011 15031
FT TRANSMEM 15032 15052
FT TRANSMEM 15053 15073
FT TRANSMEM 15074 15094
FT TRANSMEM 15095 15115
FT TRANSMEM 15116 15136
FT TRANSMEM 15137 15157
FT TRANSMEM 15158 15178
FT TRANSMEM 15179 15199
FT TRANSMEM 15200 15221
FT TRANSMEM 15222 15242
FT TRANSMEM 15243 15263
FT TRANSMEM 15264 15284
FT TRANSMEM 15285 15305
FT TRANSMEM 15306 15326
FT TRANSMEM 15327 15347
FT TRANSMEM 15348 15368
FT TRANSMEM 15369 15389
FT TRANSMEM 15390 15410
FT TRANSMEM 15411 15431
FT TRANSMEM 15432 15452
FT TRANSMEM 15453 15473
FT TRANSMEM 15474 15494
FT TRANSMEM 15495 15515
FT TRANSMEM 15516 15536
FT TRANSMEM 15537 15557
FT TRANSMEM 15558 15578
FT TRANSMEM 15579 15599
FT TRANSMEM 15600 15621
FT TRANSMEM 15622 15642
FT TRANSMEM 15643 15663
FT TRANSMEM 15664 15684
FT TRANSMEM 15685 15705
FT TRANSMEM 15706 15726
FT TRANSMEM 15727 15747
FT TRANSMEM 15748 15768
FT TRANSMEM 15769 15789
FT TRANSMEM 15790 15810
FT TRANSMEM 15811 15831
FT TRANSMEM 15832 15852
FT TRANSMEM 15853 15873
FT TRANSMEM 15874 15894
FT TRANSMEM 15895 15915
FT TRANSMEM 15916 15936
FT TRANSMEM 15937 15957
FT TRANSMEM 15958 15978
FT TRANSMEM 15979 15999
FT TRANSMEM 16000 16021
FT TRANSMEM 16022 16042
FT TRANSMEM 16043 16063
FT TRANSMEM 16064 16084
FT TRANSMEM 16085 16105
FT TRANSMEM 16106 16126
FT TRANSMEM 16127 16147
FT TRANSMEM 16148 16168
FT TRANSMEM 16169 16189
FT TRANSMEM 16190 16210
FT TRANSMEM 16211 16231
FT TRANSMEM 16232 16252
FT TRANSMEM 16253 16273
FT TRANSMEM 16274 16294
FT TRANSMEM 16295 16315
FT TRANSMEM 16316 16336
FT TRANSMEM 16337 16357
FT TRANSMEM 16358 16378
FT TRANSMEM 16379 16399
FT TRANSMEM 16400 16421
FT TRANSMEM 16422 16442
FT TRANSMEM 16443 16463
FT TRANSMEM 16464 16484
FT TRANSMEM 16485 16505
FT TRANSMEM 16506 16526
FT TRANSMEM 16527 16547
FT TRANSMEM 16548 16568
FT TRANSMEM 16569 16589
FT TRANSMEM 16590 16610
FT TRANSMEM 16611 16631
FT TRANSMEM 16632 16652
FT TRANSMEM 16653 16673
FT TRANSMEM 16674 16694
FT TRANSMEM 16695 16715
FT TRANSMEM 16716 16736
FT TRANSMEM 16737 16757
FT TRANSMEM 16758 16778
FT TRANSMEM 16779 16799
FT TRANSMEM 16800 16821
FT TRANSMEM 16822 16842
FT TRANSMEM 16843 16863
FT TRANSMEM 16864 16884
FT TRANSMEM 16885 16905
FT TRANSMEM 16906 16926
FT TRANSMEM 16927 16947
FT TRANSMEM 16948 16968
FT TRANSMEM 16969 16989
FT TRANSMEM 16990 17010
FT TRANSMEM 17011 17031
FT TRANSMEM 17032 17052
FT TRANSMEM 17053 17073
FT TRANSMEM 17074 17094
FT TRANSMEM 17095 17115
FT TRANSMEM 17116 17136
FT TRANSMEM 17137 17157
FT TRANSMEM 17158 17178
FT TRANSMEM 17179 17199
FT TRANSMEM 17200 17221
FT TRANSMEM 17222 17242
FT TRANSMEM 17243 17263
FT TRANSMEM 17264 17284
FT TRANSMEM 17285 17305
FT TRANSMEM 17306 17326
FT TRANSMEM 17327 17347
FT TRANSMEM 17348 17368
FT TRANSMEM 17369 17389
FT TRANSMEM 17390 17410
FT TRANSMEM 17411 17431
FT TRANSMEM 17432 17452
FT TRANSMEM 17453 17473
FT TRANSMEM 17474 17494
FT TRANSMEM 17495 17515
FT TRANSMEM 17516 17536
FT TRANSMEM 17537 17557
FT TRANSMEM 17558 17578
FT TRANSMEM 17579 17599
FT TRANSMEM 17600 17621
FT TRANSMEM 17622 17642
FT TRANSMEM 17643 17663
FT TRANSMEM 17664 17684
FT TRANSMEM 17685 17705
FT TRANSMEM 17706 17726
FT TRANSMEM 17727 17747
FT TRANSMEM 17748 17768
FT TRANSMEM 17769 17789
FT TRANSMEM 17790 17810
FT TRANSMEM 17811 17831
FT TRANSMEM 17832 17852
FT TRANSMEM 17853 17873
FT TRANSMEM 17874 17894
FT TRANSMEM 17895 17915
FT TRANSMEM 17916 17936
FT TRANSMEM 17937 17957
FT TRANSMEM 17958 17978
FT TRANSMEM 17979 17999
FT TRANSMEM 18000 18021
FT TRANSMEM 18022 18042
FT TRANSMEM 18043 18063
FT TRANSMEM 18064 1808
```

Search completed: April 6, 2004, 16:07:59  
Job time : 6.80374 secs

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Parotid gland;
RX MEDLINE=86033799; PubMed=3840480;
RA Clemens S., Mehansho H., Carlson D.M.;
RT "Novel multigene families encoding highly repetitive peptide
RT sequences. Sequence analyses of rat and mouse proline-rich protein
RT CDNAs.";
RL J. Biol. Chem. 260:13471-13477 (1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11901; AAA41948.1; -
DR FIR; B29149; B29149.
KW Repeat; Parotid gland; Multigene family; Signal.
FT SIGNAL.
FT CHAIN 1 16 POTENTIAL.
FT CHAIN 17 >172 ACIDIC PROLINE-RICH PROTEIN PRP25.
FT NON TER 172 172
SQ SEQUENCE 172 AA; 17416 MW; F63BFBD05459D6EA CRC64;

Query Match 75.5%; Score 40; DB 1; Length 172;
Best Local Similarity 85.7%; Pred. No. 11;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGPP 8
DB 85 GKPGQPP 91
|:|||||
|:|||||

RESULT 15
PRPP HUMAN STANDARD; PRT; 174 AA.
AC P81489;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein 11-1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RC TISSUE=Saliva;
RX MEDLINE=93385383; PubMed=8373986;
RA Kauffman D.L., Keller P.J., Bennick A., Blum M.;
RT "Alignment of amino acid and DNA sequences of human proline-rich
RT proteins.";
RL Crit. Rev. Oral Biol. Med. 4:287-292(1993).
DR GO; GO:0005576; C:extracellular; NAS.
DR InterPro; IPR000637; AT Hook.
DR PRINTS; PR00929; ATHOOK.
KW Repeat; Parotid gland; Multigene family.
FT NON TER 174 174
SQ SEQUENCE 174 AA; 17802 MW; D645F106EB1B5B5E CRC64;

Query Match 75.5%; Score 40; DB 1; Length 174;
Best Local Similarity 85.7%; Pred. No. 11;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGPP 8
DB 38 GKPGQPP 44
|:|||||
|:|||||

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 32.5514 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-5  
Perfect score: 53  
Sequence: 1 RGRPQGGP 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 25:
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mbc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_virus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	90.6	256	16 Q7UYG7	Q7UYG7 rhodopirell
2	45	84.9	93	4 Q15215	Q15215 homo sapien
3	45	84.9	173	4 Q15214	Q15214 homo sapien
4	45	84.9	188	6 Q29427	Q29427 macaca fasc
5	45	84.9	238	4 Q00600	Q00600 homo sapien
6	45	84.9	297	4 Q16038	Q16038 homo sapien
7	45	84.9	338	4 Q86YAL	Q86YAL homo sapien
8	45	84.9	358	4 Q08805	Q08805 homo sapien
9	45	84.9	382	4 Q00599	Q00599 homo sapien
10	43	81.1	188	11 Q62106	Q62106 mus musculus
11	43	81.1	295	11 Q07611	Q07611 rattus norv
12	43	81.1	317	11 Q62103	Q62103 mus musculus
13	43	81.1	803	16 Q812W2	Q812W2 streptomyce
14	42	79.2	85	5 Q3VR11	Q3VR11 drosophila
15	42	79.2	109	3 Q9P8A8	Q9P8A8 agaricus bi
16	42	79.2	170	11 Q07610	Q07610 rattus norv

17	79.2	204	16	Q7U416	Q7U416 synchococc
18	79.2	339	13	Q7ZUX2	Q7ZUX2 brachydanio
19	79.2	796	3	Q8NKJ3	Q8NKJ3 saccharomyc
20	77.4	220	13	Q7T339	Q7T339 brachydanio
21	75.5	71	15	Q8AIV8	Q8AIV8 human immun
22	75.5	147	11	Q62104	Q62104 mus musculus
23	75.5	164	11	Q04105	Q04105 rattus norv
24	75.5	202	11	Q04117	Q04117 rattus norv
25	75.5	274	11	Q04154	Q04154 rattus norv
26	75.5	309	4	Q04118	Q04118 homo sapien
27	75.5	325	3	Q59748	Q59748 schizosacch
28	75.5	395	11	Q9JIB1	Q9JIB1 rattus norv
29	75.5	683	16	Q7UR32	Q7UR32 rhodopirell
30	75.5	1559	5	Q86BS0	Q86BS0 drosophila
31	75.5	1569	5	Q9VF66	Q9VF66 drosophila
32	75.5	1622	5	Q86BP0	Q86BP0 drosophila
33	75.5	1736	4	Q75111	Q75111 homo sapien
34	75.5	1857	4	Q95153	Q95153 homo sapien
35	73.6	46	4	Q15218	Q15218 homo sapien
36	73.6	99	15	Q7ZB17	Q7ZB17 chimpanzee
37	73.6	209	4	Q8N811	Q8N811 homo sapien
38	73.6	266	10	Q40857	Q40857 picea glauc
39	73.6	300	11	Q61888	Q61888 mus musculu
40	73.6	504	11	Q91X93	Q91X93 mus musculu
41	73.6	584	4	Q9NXG8	Q9NXG8 homo sapien
42	73.6	591	13	P79817	P79817 oryzias lat
43	73.6	662	4	Q9H8X6	Q9H8X6 homo sapien
44	73.6	666	13	Q8JHT8	Q8JHT8 oryzias lat
45	73.6	670	13	Q8J122	Q8J122 oryzias lat

## ALIGNMENTS

RESULT 1				
ID	Q7UYG7	PRELIMINARY;	PRT; 256 AA.	
AC	Q7UYG7;	01-OCT-2003 (T-EMBLrel. 25, Created)		
DT	01-OCT-2003 (T-EMBLrel. 25, Last sequence update)			
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)			
DE	Hypothetical protein.			
GN	RB624.			
OS	Rhodopirellula baltica.			
OC	Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;			
OC	Planctomycetaceae; Pirellula.			
OX	NCBI_taxid=117;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=1;			
RX	MEDLINE=22735913; PubMed=12835416;			
RA	Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,			
RA	Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,			
RA	Schlesner H., Amann R., Reinhardt R.;			
RT	"Complete genome sequence of the marine planctomycete Pirellula sp.			
RT	strain 1.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).			
DR	EMBL; BX294134; CAD71675.1; ..			
KW	Hypothetical protein; Complete proteome.			
QW	SEQUENCE 256 AA; 27003 MW; 2FAA06B5206E8B2F CRC64;			
Query Match				90.6%; Score 48; DB 16; Length 256;
Best Local Similarity				100.0%; Pred. No. 1.9;
Matches 8; Conservative				0; Mismatches 0; Indels 0; Gaps 0;
QY	1 RGRPQGGP 8			
DB	160 RGRPQGGP 167			
RESULT 2				
ID	Q15215	PRELIMINARY;	PRT; 93 AA.	

AC Q15215;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Salivary proline-rich protein 1 (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84298176; PubMed=6089212;  
RA Azen E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S.,  
RA Maeda N., Vanin E.F., Carlson D.M., Smithies O.,  
RT "Clones from the human gene complex coding for salivary proline-rich  
proteins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565 (1984).  
DR EMBL; K02576; AAA36503.1; -;  
FT NON\_TER 1  
SQ SEQUENCE 93 AA; 9170 MW; 04AD5F0797E31867 CRC64;  
  
Query Match 84.9%; Score 45; DB 4; Length 93;  
Best Local Similarity 87.5%; Pred. No. 2.1;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 GRPQGPPO 9  
Db 41 GKPQGPPO 48  
  
RESULT 3  
Q15214  
ID Q15214 PRELIMINARY; PRT; 173 AA.  
AC Q15214;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Salivary proline-rich protein 1 (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84298176; PubMed=6089212;  
RA Azen E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S.,  
RA Maeda N., Vanin E.F., Carlson D.M., Smithies O.,  
RT "Clones from the human gene complex coding for salivary proline-rich  
proteins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565 (1984).  
DR EMBL; K02575; AAA36502.1; -;  
GO; GO:0005576; C:extracellular; NAS.  
FT NON\_TER 1  
FT NON\_TER 173 173  
SQ SEQUENCE 173 AA; 17206 MW; BB4E3035B77A7C0 CRC64;  
  
Query Match 84.9%; Score 45; DB 4; Length 173;  
Best Local Similarity 87.5%; Pred. No. 4;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 GRPQGPPO 9  
Db 160 GKPQGPPO 167  
  
RESULT 4  
Q29427  
ID Q29427 PRELIMINARY; PRT; 188 AA.  
AC Q29427;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Proline-rich protein.

GN MNP4 OR MNP4P9.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92009216; PubMed=1916292;  
RA Lin H.H., Kousvelari E.E., Ann D.K.,  
RT "Sequence and expression of the Mnp4 gene encoding basic proline-rich  
protein in macaque salivary glands.";  
RL Gene 104:219-226 (1991).  
DR EMBL; M81322; AAA36905.1; -;  
DR EMBL; M81321; AAA36904.1; -;  
DR PIR; JH0481; JH0481.  
SQ SEQUENCE 188 AA; 19135 MW; C8B6D0B7F4DE504 CRC64;  
  
Query Match 84.9%; Score 45; DB 6; Length 188;  
Best Local Similarity 87.5%; Pred. No. 4.4;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 GRPQGPPO 9  
Db 95 GKPQGPPO 102  
  
RESULT 5  
Q00600  
ID Q00600 PRELIMINARY; PRT; 238 AA.  
AC Q00600;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Parotid 'o' protein (Fragment).  
GN PRB4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96108975; PubMed=8554050;  
RA Azen E.A., Amberger E., Fisher S., Prakobphol A., Niece R.L.,  
RT "PRB1, PRB2, and PRB4 coded polymorphisms among human salivary  
concanavalin-A binding, 11-1, and P6 proline-rich proteins.";  
RL Am. J. Hum. Genet. 58:143-153 (1996).  
DR EMBL; S80916; AAB50687.2; -;  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 238 AA; 24005 MW; 4F44E947FFF3A6C1 CRC64;  
  
Query Match 84.9%; Score 45; DB 4; Length 238;  
Best Local Similarity 87.5%; Pred. No. 5.6;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 GRPQGPPO 9  
Db 172 GKPQGPPO 179  
  
RESULT 6  
Q16038  
ID Q16038 PRELIMINARY; PRT; 297 AA.  
AC Q16038;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE PRBIM protein (Fragment).  
GN PRBIM.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;

```

RN  SEQUENCE FROM N.A.
RP  MEDLINE=93304421; PubMed=8317492;
RX  Azen E.A., Latreille P., Niece R.L.;
RA  "PRBI gene variants coding for length and null polymorphisms among
RT  human salivary Ps, PmF, PmS, and Pe proline-rich proteins (PRPs).";
RL  Am. J. Hum. Genet. 53:264-278(1993).
DR  EMBL; S62941; AAB27288.2; -.
DR  PIR; B40750; FIHUB6.
FT  PIR; D40750; D40750.
FT  NON TER
SQ  SEQUENCE 297 AA; 29046 MW; 1C7BFE4CA6B5B5F0 CRC64;

Query Match      84.9%; Score 45; DB 4; Length 297;
Best Local Similarity 87.5%; Pred. No. 7;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY  2 GRPQGPQ 9
Db  143 GKPGPPQ 150

RESULT 7
ID  Q86YA1 PRELIMINARY; PRT; 338 AA.
AC  Q86YA1;
DT  01-JUN-2003 (TrEMBLrel. 24, Created)
DT  01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Hypothetical protein (Fragment).
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Skin;
RA  Strausberg R.;
RL  Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR  EMBL; BC044827; AAH4827.1; -.
KW  Hypothetical protein.
FT  NON TER
SQ  SEQUENCE 338 AA; 33344 MW; 7F54B4E5AB002261 CRC64;

Query Match      84.9%; Score 45; DB 4; Length 338;
Best Local Similarity 87.5%; Pred. No. 8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY  2 GRPQGPQ 9
Db  184 GKPGPPQ 191

RESULT 8
ID  Q08805 PRELIMINARY; PRT; 358 AA.
AC  Q08805;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  PRB1 protein (Fragment).
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93304421; PubMed=8317492;
RA  Azen E.A., Latreille P., Niece R.L.;
RT  "PRBI gene variants coding for length and null polymorphisms among
RT  human salivary Ps, PmF, PmS, and Pe proline-rich proteins (PRPs).";
RL  Am. J. Hum. Genet. 53:264-278(1993).

```

```

DR  EMBL; S62941; AAB27289.1; -.
DR  PIR; D40750; D40750.
FT  NON TER
SQ  SEQUENCE 358 AA; 35050 MW; DB7F87B8D5EA759E CRC64;

Query Match      84.9%; Score 45; DB 4; Length 358;
Best Local Similarity 87.5%; Pred. No. 8.4;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY  2 GRPQGPQ 9
Db  204 GKPGPPQ 211

RESULT 9
ID  Q00599 PRELIMINARY; PRT; 382 AA.
AC  Q00599;
DT  01-JUL-1997 (TrEMBLrel. 04, Created)
DT  01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Con1 (Fragment).
DE  Con1 (Fragment).
GN  PRB2.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=96108975; PubMed=8554050;
RA  Azen E.A., Amberger E., Fisher S., Prakobphol A., Niece R.L.;
RT  "PRB1, PRB2, and PRB4 coded polymorphisms among human salivary
RT  concanavalin-A binding, II-1, and Pe proline-rich proteins.";
RL  Am. J. Hum. Genet. 58:143-153(1996).
DR  EMBL; S80905; AAB50686.1; -.
FT  NON TER
SQ  SEQUENCE 382 AA; 37277 MW; D500942AB979209D CRC64;

Query Match      84.9%; Score 45; DB 4; Length 382;
Best Local Similarity 87.5%; Pred. No. 9;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY  2 GRPQGPQ 9
Db  330 GKPGPPQ 337

RESULT 10
ID  Q62106 PRELIMINARY; PRT; 188 AA.
AC  Q62106;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Proline-rich salivary protein (Fragment).
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=86033799; PubMed=3840480;
RA  Clements S., Mehansho H., Carlson D.M.;
RT  "Novel multigene families encoding highly repetitive peptide
RT  sequences: Sequence analyses of rat and mouse proline-rich protein
RT  cDNAs.";
RL  J. Biol. Chem. 260:13471-13477(1985).
DR  EMBL; W19413; AAA40002.1; -.
DR  PIR; D29149; D29149.
FT  NON TER
SQ  SEQUENCE 188 AA; 18657 MW; A6B98BC5560473FE CRC64;

Query Match      81.1%; Score 43; DB 11; Length 188;

```

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Best Local Similarity 87.5%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPQ 9  
| | | | |  
Db 303 GNPQGPQ 310

QY 2 GRPQGPQ 9  
| | | | |  
Db 174 GNPQGPQ 181

RESULT 13  
Q81ZW2 PRELIMINARY; PRT; 803 AA.  
AC Q81ZW2;  
DT 01-JUN-2003 (TREMELrel. 24, Created)  
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Putative serine/threonine protein kinase.  
GN FAN27 OR SAV5424.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
RT metabolites.";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.;  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis.";  
RL Nat. Biotechnol. 21:526-531(2003).  
DR GO:0005524; F:ATP binding; IEA.  
DR GO:0004574; F:protein serine/threonine kinase activity; IEA.  
DR GO:0004713; F:protein-tyrosine kinase activity; IEA.  
DR GO:0006118; P:electron transport; IEA.  
DR GO:0006468; P:protein amino acid phosphorylation; IEA.  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; Pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TK; 1.  
DR SMART; SM00219; TyKc; 1.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_ST; 1.  
DR Kinase; Serine/threonine-protein kinase; Complete proteome.  
SQ SEQUENCE 803 AA; 82197 MW; C38C2B50737FDB66 CRC64;

Query Match 81.1%; Score 43; DB 16; Length 803;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGPQ 9  
| | | | |  
Db 273 GNPQGPQ 280

RESULT 12  
Q62103 PRELIMINARY; PRT; 317 AA.  
AC Q62103;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Proline-rich protein precursor.  
GN PRP2 OR PRP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD-1; TISSUE=Liver;  
RX MEDLINE=88273214; PubMed=2839509;  
RA Ann D.K., Smith M.K., Carlson D.M.;  
RT "Molecular evolution of the mouse proline-rich protein multigene  
RT family. Insertion of a long interspersed repeated DNA element.";  
RL J. Biol. Chem. 263:10887-10893(1998).  
DR EMBL; M23236; AAA53048.1; -.  
DR FIR; A28996; A28996.  
DR MGD; MGI:1932491; Prp2.  
KW Signal.  
FT SIGNAL.  
SQ CHAIN 16 317 PROLINE-RICH PROTEIN  
SEQUENCE 317 AA; 31719 MW; 019301BE31D73278 CRC64;

Query Match 81.1%; Score 43; DB 11; Length 317;  
Best Local Similarity 87.5%; Pred. No. 16;

QY 2 GRPQGPQ 9  
| | | | |  
Db 195 GNPQGPQ 201

RESULT 14  
Q9VR11

Query Match 81.1%; Score 43; DB 11; Length 317;  
Best Local Similarity 87.5%; Pred. No. 16;

```

ID Q9VR11 PRELIMINARY; PRT; 85 AA.
AC Q9VR11;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CG1678 protein.
DE CG1678.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Stapling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.C.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RL EMBL; AE003568; AAP50815.1; -.
DR FlyBase; FBgn031176; CG1678.
SQ SEQUENCE 85 AA; 8360 MW; 0F04923B4AEB66CE CRC64;

Query Match 79.2%; Score 42; DB 5; Length 85;
Best Local Similarity 77.8%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RGRPQGPQ 9
Db 52 QGLPQGPQ 60

RESULT 15
Q9P8A8 PRELIMINARY; PRT; 109 AA.
ID Q9P8A8
AC Q9P8A8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN DAG8.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Horst U3;
RA Eastwood D.C., Kingsnorth C.S., Jones H., Burton K.S.;
RT "Development of Agaricus bisporus sporophores following tissue
RT detachment is maintained by the expression of stress tolerance and
RT nutritional genes.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ271699; CAB85692.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 109
FT NON_TER 109
SQ SEQUENCE 109 AA; 11293 MW; FF4CB01CF2BD2ED4 CRC64;

Query Match 79.2%; Score 42; DB 3; Length 109;
Best Local Similarity 77.8%; Pred. No. 8.1;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RGRPQGPQ 9
Db 43 QGLPQGPQ 51

Search completed: April 6, 2004, 16:14:39
Job time : 32.6283 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 50.3832 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-5  
Perfect score: 53  
Sequence: 1 RGRPGGPPQ 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	9	4 AAB48775	Aab48775 Human sal
2	53	100.0	10	4 AAB48776	Aab48776 Human sal
3	53	100.0	17	4 AAB48783	Aab48783 Human sal
4	53	100.0	132	4 AAB38848	Abb38848 Peptide #
5	53	100.0	132	4 AAM32323	Aam32323 Peptide #
6	53	100.0	132	4 AAM72058	Aam72058 Human bon
7	53	100.0	132	4 AAM59494	Aam59494 Human bra
8	53	100.0	132	4 ABG53742	Abg53742 Human liv
9	53	100.0	132	5 ABG41873	Abg41873 Human pep
10	53	100.0	149	6 ABR57423	Abr57423 Human NOV
11	53	100.0	154	6 ABR56769	Abr56769 Human sec
12	53	100.0	166	6 ADA83798	Ada83798 Human PRH
13	53	100.0	166	7 ADC98216	Adc98216 Human sal
14	48	90.6	8	4 AAB48774	Aab48774 Human sal
15	48	90.6	44	2 AAW03557	Aaw03557 Human pro
16	48	90.6	45	1 AAP93320	Aap93320 P-C gene.
17	48	90.6	111	6 ABP75970	Abp75970 Human GEN
18	48	90.6	111	6 ABP76136	Abp76136 Human GEN
19	46	86.8	106	6 ABP76137	Abp76137 Human GEN
20	46	86.8	106	6 ABP75971	Abp75971 Human GEN
21	45	84.9	117	6 AAO30249	Aao30249 Human Sap
22	45	84.9	192	6 AAO30246	Aao30246 Human Sap
23	45	84.9	208	6 AAO30188	Aao30188 Human CP3
24	45	84.9	247	6 ABU11891	Abu11891 Human ABC
25	45	84.9	331	7 ADD45740	Add45740 Human Pro

26 43 81.1 162 7 AAO30407 Aao30407 Human sec  
27 43 81.1 267 7 AAO30408 Aao30408 Human sec  
28 43 81.1 296 7 ADD45442 Add45442 Rat Prote  
29 43 81.1 296 7 ADE57105 Ade57105 Rat Prote  
30 43 81.1 591 4 ABG15767 Abg15767 Novel hum  
31 43 81.1 591 4 ABG18110 Abg18110 Novel hum  
32 42 79.2 85 4 ABB58372 Abb58372 Drosophil  
33 42 79.2 796 6 ABR53413 Abr53413 Protein s  
34 41 77.4 7 4 AAB48773 Aab48773 Human sal  
35 41 77.4 1146 6 AAE34447 Aae34447 Human lip  
36 40 75.5 18 2 AAW03556 Aaw03556 Octadecap  
37 40 75.5 41 6 AAO30248 Aao30248 Human Sap  
38 40 75.5 49 3 AAB34155 Aab34155 Gene 13 h  
39 40 75.5 80 3 AAY65169 Aay65169 Human 5,  
40 40 75.5 82 2 AAW50193 Aaw50193 Amino aci  
41 40 75.5 124 2 AAW50192 Aaw50192 Amino aci  
42 40 75.5 164 2 ADD47283 Add47283 Rat Prote  
43 40 75.5 202 7 ADD48701 Add48701 Rat Prote  
44 40 75.5 266 4 ABG05763 Abg05763 Novel hum  
45 40 75.5 274 7 ADD47240 Add47240 Rat Prote

## ALIGNMENTS

RESULT 1  
AAB48775  
ID AAB48775 standard; peptide; 9 AA.  
XX  
AC AAB48775;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human saliva PRP-1 fragment (residues 106-114), SEQ ID NO:5.  
XX  
KW Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW Chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX  
OS Homo sapiens.  
XX  
PN WO200069890-A1.  
XX  
PD 23-NOV-2000.  
XX  
PF 11-MAY-2000; 2000WO-SE000930.  
XX  
PR 17-MAY-1999; 99SE-00001773.  
XX  
PA (STRO/) STROEMBERG N.  
PA (JOHA/) JOHANSSON I.  
XX  
PI Stroemberg N, Johansson I;  
XX  
DR WPI; 2001-031923/04.  
XX  
PT New oligopeptides comprising 2 arginine residues from degradation of  
PT proline-rich proteins, useful for preventing dental caries.  
XX  
PS Claim 4; Page 24; 36pp; English.  
XX  
CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
CC AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

CC derived oligopeptides of the invention

SQ Sequence 9 AA;

Query Match 100.0%; Score 53; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQGPPQ 9  
| | | | |  
Db 1 RGRPQGPPQ 9

RESULT 2

AA848776  
ID AAB48776 standard; peptide; 10 AA.

AC AAB48776;

XX 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 106-115), SEQ ID NO:6.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.

XX Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SR000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of  
PT proline-rich proteins, useful for preventing dental caries.

PS Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
CC AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
CC derived oligopeptides of the invention

XX Sequence 10 AA;

Query Match 100.0%; Score 53; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQGPPQ 9  
| | | | |  
Db 1 RGRPQGPPQ 9

RESULT 3

AA848783  
ID AAB48783 standard; peptide; 17 AA.

XX AAB48783;

XX 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.

XX Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SR000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04;

XX New oligopeptides comprising 2 arginine residues from degradation of  
PT proline-rich proteins, useful for preventing dental caries.

PS Claim 2; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
CC AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
CC derived oligopeptides of the invention

XX Sequence 17 AA;

Query Match 100.0%; Score 53; DB 4; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPQGPPQ 9  
| | | | |  
Db 8 RGRPQGPPQ 16

RESULT 4

ABB38848  
ID ABB38848 standard; peptide; 132 AA.

XX ABB38848;

XX 04-FEB-2002 (first entry)

XX Peptide #6354 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

```

XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX PS Claim 27; SEQ ID NO 31483; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human foetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 132 AA;
XX
Query Match 100.0%; Score 53; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRPGGPPQ 9
Db 88 RGRPGGPPQ 96
RESULT 5
AAM32323
ID AAM32323 standard; protein; 132 AA.
XX AC AAM32323;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #6360 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX OS Genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.

```

```

PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX PS Claim 27; SEQ ID NO 32592; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX SQ Sequence 132 AA;
XX
Query Match 100.0%; Score 53; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRPGGPPQ 9
Db 88 RGRPGGPPQ 96
RESULT 6
AAM72058
ID AAM72058 standard; protein; 132 AA.
XX AC AAM72058;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32364.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX PS Example 4; SEQ ID NO 32364; 658pp + Sequence Listing; English.
XX

```

CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention  
 XX Sequence 132 AA;  
 SQ

Query Match 100.0%; Score 53; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGPQ 9  
 DB 88 RGRPQGPQ 96

RESULT 7  
 AAM59494  
 ID AAM59494 standard; protein; 132 AA.

XX AC AAM59494;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31599.

XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.

XX OS Homo sapiens.

XX PN WO200157275-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000667.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains.  
 XX Example 4; SEQ ID NO 31599; 650pp + Sequence Listing; English.

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 53; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGPQ 9

DB 88 RGRPQGPQ 96

RESULT 8  
 ABG53742

ID ABG53742 standard; peptide; 132 AA.

XX AC ABG53742;

XX DT 25-FEB-2003 (first entry)

XX DE Human liver peptide, SEQ ID NO 32390.

XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.

XX OS Homo sapiens.

XX PN WO200157273-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000664.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488898/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human adult liver.  
 XX Claim 27; SEQ ID NO 32390; 658pp; English.

XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
 CC liver single exon encoded peptides of the invention. Note: The sequence  
 CC information for this patent does not appear in the printed specification  
 CC but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 53; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGPQ 9  
 DB 88 RGRPQGPQ 96

RESULT 9  
 ABG41873

ID ABG41873 standard; peptide; 132 AA.

CC tuberosus sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 132 AA;  
Query Match 100.0%; Score 53; DB 5; Length 132;  
Best Local Similarity 100.0%; Pred. No. 1,2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 RGRPGGPPQ 9  
Db 88 RGRPGGPPQ 96  
RESULT 10  
ABR57423  
ID ABR57423 standard; protein; 149 AA.  
XX AC ABR57423;  
XX DT 15-SEP-2003 (first entry)  
XX DE Human NOV7 protein SEQ ID NO:14.  
XX KW Human; NOVX; cytostatic; cardiant; antiinflammatory; immunosuppressive;  
KW anti-allergic; haemostatic; anti-HIV; antidiabetic; antiarteriosclerotic;  
KW anorectic; antiaesthetic; nephrotropic; antiarthritic; hepatotropic;  
KW neuroprotective; nontropic; antibacterial; virucide; antiparasitic;  
KW relaxant; anticonvulsant; hypotensive; vasotropic; vaccine; cancer;  
KW vulnerary; angiogenic; antiangiogenic; gene therapy; inflammation;  
KW cardiomyopathy; atherosclerosis; hypertension; diabetes; asthma;  
KW autoimmune disorder; allergy; blood disorder; AIDS; obesity; stroke;  
KW acquired immunodeficiency syndrome; nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; Parkinson's disease; Goitre; infection; stroke;  
KW muscular dystrophy; epilepsy; wasting disorder; chromosome 12.  
XX OS Homo sapiens.  
XX FN WO200294870-A2.  
XX PD 28-NOV-2002.  
XX PF 02-NOV-2001; 2001WO-US051580.  
XX PR 02-NOV-2000; 2000US-0245291P.  
XX PR 02-NOV-2000; 2000US-0245317P.  
XX PR 07-NOV-2000; 2000US-0246562P.  
XX PR 08-NOV-2000; 2000US-0246871P.  
XX PR 26-JAN-2001; 2001US-0264389P.  
XX PR 26-JAN-2001; 2001US-0264423P.  
XX PR 29-JAN-2001; 2001US-0264799P.  
XX PA (CURA-) CURAGEN CORP.  
XX PI Grosse WM, Macdougall JR, Smithson G, Millet I, Stone DJ;  
PI Gunther E, Ellerman K, Alsobrook JP, Lepley DM, Burgess CB, Li L;  
PI Spyttek KA, Edinger SR, Gangoli EA, Gorman L, Taupier RJ, Li L;  
PI Qio X, Fernandes ER, Vernet CM, Tchernev VT, Casman SD, Shenoy S;  
PI Mishra V, Furtak K, Baugartner JC, Colman SD;  
XX WPI; 2003-140359/13.  
XX DR N-PSDB; ACF03558.  
XX PT New NOVX polypeptide useful for preventing or treating NOVX-associated  
XX disorders, e.g. cancer, cardiomyopathy, atherosclerosis or diabetes, and  
XX in chromosome mapping, tissue typing or pharmacogenomics.

ABG41873;  
19-AUG-2002 (first entry)  
Human peptide encoded by genome-derived single exon probe SEQ ID 31538.  
Human; single exon probe; asthma, lung cancer, COPD; ILD;  
chronic obstructive pulmonary disease; interstitial lung disease;  
familial idiopathic pulmonary fibrosis; neurofibromatosis;  
tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;  
Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
pulmonary histiocytosis; lymphangiomyomatosis; Karagener syndrome;  
pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
primary ciliary dyskinesia; pulmonary hypertension;  
hyaline membrane disease.  
Homo sapiens.  
WO200186003-A2.  
15-NOV-2001;  
30-JAN-2001; 2001WO-US000665.  
04-FEB-2000; 2000US-0180312P.  
26-MAY-2000; 2000US-0207456P.  
30-JUN-2000; 2000US-00608408.  
03-AUG-2000; 2000US-00632366.  
21-SEP-2000; 2000US-0234687P.  
27-SEP-2000; 2000US-0236359P.  
04-OCT-2000; 2000GB-00024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2002-114183/15.  
Spatially-addressable set of single exon nucleic acid probes, used to  
measure gene expression in human lung samples.  
Claim 27; SEQ ID NO 31538; 634pp; English.  
The invention relates to a spatially-addressable set of single exon  
nucleic acid probes for measuring gene expression in a sample derived  
from human lung comprising single exon nucleic acid probes having one of  
12614 nucleic acid sequences mentioned in the specification, or their  
complements or the 12387 open reading frames derived from the 12614  
probes. Also included are a microarray comprising the novel set of probes  
; the novel set of probes which hybridise at high stringency to a nucleic  
acid expressed in the human lung; measuring gene expression in a sample  
derived from human lung, comprising (a) contacting the array with a  
collection of detectably labeled nucleic acids derived from human lung  
mRNA, and (b) measuring the label detectably bound to each probe of the  
array; identifying exons in a eukaryotic genome, comprising (a)  
algorithmically predicting at least one exon from genomic sequences of  
the eukaryote; and (b) detecting specific hybridisation of detectably  
labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
having a fragment identical to the predicted exon; the probe is included  
in the above mentioned microarray; assigning exons to a single gene,  
comprising (a) identifying exons from genomic sequence by the method  
above and (b) measuring the expression of each of the exons in several  
tissues and/or cell types using hybridisation to a common pattern of  
microarrays having a probe with the exon, where a common pattern of  
expression of the exons in the tissues and/or cell types indicates that  
the exons should be assigned to a single gene; a peptide comprising one  
of 12011 sequences, mentioned in the specification, or encoded by the  
probes/open reading frames (ORF). The probes are used for gene expression  
analysis, and for identifying exons in a gene, particularly using human  
lung derived mRNA and for the study of lung diseases such as asthma, lung  
cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,

XX PS Claim 1; Page 69; 346pp; English.

XX CC ACF03547 to ACF03570 encode the human NOVX proteins (I) given in ABR5741.2

XX CC to ABR57435. (I) have cytostatic, cardiant, antiinflammatory, neurotropic,

XX CC immunosuppressive, anti-allergic, haemostatic, anti-HIV, antidiabetic,

XX CC antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, virucide,

XX CC antiarthritic, hepatotropic, neuroprotective, antibacterial, relaxant,

XX CC antiparasitic, anticonvulsant, hypotensive, vasotropic, antiparkinsonian,

XX CC vulnary, angiogenic and angiogenic activities, and can be used in

XX CC gene therapy and vaccines. The NOVX polypeptides and their antibodies can

XX CC be used to determine the presence or absence of (I) in a sample. The NOVX

XX CC polypeptides, polynucleotides encoding them, and antibodies against them,

XX CC are useful in manufacturing a medicament for treating or preventing a

XX CC syndrome associated with a NOVX-associated disorder such as hypertension,

XX CC cardiomyopathy, atherosclerosis, cancer, diabetes, asthma, inflammation,

XX CC autoimmune disorders, allergies, blood disorders, obesity, acquired

XX CC immunodeficiency syndrome (AIDS), immunoglobulin (IgA) nephropathy,

XX CC cirrhosis, arthritis, Alzheimer's disease, Parkinson's disease, goitre,

XX CC infections (e.g. bacterial, viral, parasitic), stroke, muscular

XX CC dystrophy, epilepsy, and other wasting disorders associated with chronic

XX CC diseases. ACF03571 to ACF03644 represent PCR primers and probes for NOVX

XX CC sequence, which are used in an example from the present invention.

XX SQ Sequence 149 AA;

Query Match 100.0%; Score 53; DB 6; Length 149;

Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPGPPQ 9

Db 105 RGRPGPPQ 113

RESULT 11

ABR56769

ID ABR56769 standard; protein; 154 AA.

XX AC ABR56769;

XX DT 30-JUL-2003 (first entry)

DE Human secreted protein SSCP-44 SEQ ID NO:44.

XX KW Human; secreted protein; SSCP; cytostatic; antiarteriosclerotic;

XX KW anticonvulsant; neurotropic; neuroprotective; cerebroprotective; anti-HIV;

XX KW anti-allergic; antiinflammatory; thyromimetic; gene therapy; cancer;

XX KW cell proliferative disorder; atherosclerosis; neurological disorder;

XX KW epilepsy; Huntington's disease; stroke; immune disorder; allergy; AIDS;

XX KW inflammatory disorder; developmental disorder; hypothyroidism;

XX KW Cushing's syndrome; infection.

XX OS Homo sapiens.

XX PN WO2003016506-A2.

XX PD 27-FEB-2003.

XX PF 15-AUG-2002; 2002WO-US027143.

XX PR 17-AUG-2001; 2001US-0313249P.

XX PR 24-AUG-2001; 2001US-0314752P.

XX PR 07-SEP-2001; 2001US-0317818P.

XX PR 07-SEP-2001; 2001US-0317824P.

XX PR 21-SEP-2001; 2001US-0324040P.

XX PR 24-SEP-2001; 2001US-0324586P.

XX PR 02-NOV-2001; 2001US-0343980P.

XX PR 28-NOV-2001; 2001US-0343229P.

XX PR 13-FEB-2002; 2002US-0357002P.

XX PR 06-MAR-2002; 2002US-0362439P.

XX PR 19-MAR-2002; 2002US-0366041P.

XX PR 30-APR-2002; 2002US-0376988P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Tang TY, Warren BA, Gietzen KJ, Lal PG, Yue H, Honchell CD;

PI Lehr-Mason PM, Burford N, Xu Y, Baughn MR, Duggan EM, Tran UK;

PI Lee EA, Forsythe IU, Richardson TW, Lee S, Thangaveilu K, Yue H;

PI Emerling BM, Walia NK, Azimzai Y, Sanjwal B, Hafalia AUA;

PI Borowsky ML, Nguyen DB, Ison CH, Astronoff A, Ding L, Lee SY;

PI Becha SD, Ramkumar J, Gandhi AR, Jin P, Fu GK, Swarnakar A;

XX WPI: 2003-278569/27.

DR N-PSDB; ACC79069.

XX New human secreted proteins (SECP), useful for diagnosing, treating and

XX PT preventing diseases or conditions associated with the aberrant SSCP

XX PT expression e.g. cancer, AIDS, atherosclerosis, Huntington's disease,

XX PT stroke, infections.

XX Claim 1; Page 22; 286pp; English.

PS ACF79026 to ACF79105 encode the human secreted proteins (I) given in

XX CC ABR56726 to ABR56805, designated SSCP-1 to SSCP-80. SSCP sequences can

XX CC have cytostatic, antiarteriosclerotic, anticonvulsant, antiinflammatory,

XX CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, anti-allergic and

XX CC thymimetic activities, and can be used in gene therapy. The SSCP

XX CC proteins and polynucleotides can be used in diagnosing, treating and

XX CC preventing diseases or conditions associated with the decreased

XX CC expression or overexpression of SSCP, such as cell proliferative (e.g.

XX CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's

XX CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and

XX CC developmental (e.g. hypothyroidism, Cushing's syndrome) disorders, or

XX CC infections. They are also useful in assessing the effects of exogenous

XX CC compounds on the expression of nucleic acid and amino acid sequences of

XX CC SSCP. The SSCP or its fragments are useful in screening compounds for

XX CC effectiveness as agonist or antagonist of the polypeptides, or in

XX CC altering the expression of the target polynucleotide and compounds that

XX CC specifically bind to or modulate the activity of the polypeptide

XX SQ Sequence 154 AA;

Query Match 100.0%; Score 53; DB 6; Length 154;

Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPGPPQ 9

Db 110 RGRPGPPQ 118

RESULT 12

ADA83798

ID ADA83798 standard; protein; 166 AA.

XX AC ADA83798;

XX DT 20-NOV-2003 (first entry)

XX DE Human PRH2 protein.

XX KW human; marker; expressed sequence tag; EST; arabidopsis; tumour;

XX KW stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;

XX KW vaccine.

XX OS Homo sapiens.

XX PN WO2002103028-A2.

XX PD 27-DEC-2002.

XX PF 30-MAY-2002; 2002WO-IB004189.

XX PR 30-MAY-2001; 2001US-0293999P.

XX PR 22-OCT-2001; 2001US-0330457P.

PR 19-FEB-2002; 2002US-0357144P.

XX PA (BIOM-) BIOMEDICAL CENT.

XX Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;

XX DR WPI; 2003-175241/17.

XX DR N-PSDB; ADA83797.

XX PT Determining if a nucleic acid is a marker for a phenotype/cell type of  
PT interest, by global comparison of expressed sequence tags known to be  
PT expressed in the phenotype/cell type with all ESTs expressed in normal  
PT tissue.

XX PS Claim 29; Page 191-192; 516pp; English.

CC The invention relates to a novel method for determining if a nucleic acid  
CC is a marker for a predetermined phenotype/cell type of interest from a  
CC biological species. The method comprises performing a global comparison  
CC of a group of expressed sequence tags (ESTs) known to be expressed in the  
CC phenotype/cell type of interest with all ESTs expressed in normal tissue  
CC in order to identify ESTs that are preferentially expressed in the  
CC phenotype/cell of interest. A method of the invention is useful for  
CC determining whether a nucleic acid is a marker for a predetermined  
CC phenotype or cell type of interest from a biological species, preferably  
CC Arabidopsis or human. The cell type of interest is an abnormal cell such  
CC as a tumour cell, and the predetermined phenotype is a stress-induced  
CC phenotype such as hyperosmotic stress or high salt conditions. A method  
CC of the invention is also useful for determining the progression of colon  
CC cancer in a human, for detecting a tumour cell, and for regulating or  
CC preventing the growth of a tumour cell. An antibody of the invention is  
CC useful for detecting the absence or presence of peptides encoded by  
CC tumour-associated markers. A polypeptide of the invention is useful as an  
CC immunogen for vaccinating an animal. The present sequence represents a  
CC tumour-associated antigen of the invention.

XX SQ Sequence 166 AA;

Query Match 100.0%; Score 53; DB 6; Length 166;

Best Local Similarity 100.0%; Pred. NO. 1.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRFQGPPQ 9

DB 122 RGRFQGPPQ 130

RESULT 13

ADC98216  
ID ADC98216 standard; protein; 166 AA.

XX AC ADC98216;

XX DT 01-JAN-2004 (first entry)

XX DE Human salivary acidic proline-rich phosphoprotein (PRP).

XX KW Human; salivary acidic proline-rich phosphoprotein; PRP; autoantigen;  
KW atopic dermatitis-inducing protein; salivary gland; IgE autoantibody;  
KW immunoglobulin E; mast cell activation; basophil activation; diagnosis;  
KW risk assessment; sensitisation remedy; dermatological; antiallergic;  
KW antiinflammatory.

XX OS Homo sapiens.

XX PN WO2003084991-A1.

XX PD 16-OCT-2003.

XX PF 04-APR-2003; 2003WO-JF004325.

XX PR 08-APR-2002; 2002JP-00105425.

PA (MISC-) JAPAN SCI & TECHNOLOGY CORP.

XX Hide M, Yamamoto S, Tanaka T, Koro O;

XX DR WPI; 2003-833567/77.

XX DR N-PSDB; ADC98215.

XX PT Atopic dermatitis-inducing proteins, applicable in diagnosis of including  
PT risk of onset, and in developing sensitization remedies for the disease.

XX PS Claim 4; SEQ ID NO 2; 43pp; Japanese.

CC The invention relates to the human atopic dermatitis-inducing proteins,  
CC salivary acidic proline-rich phosphoprotein (PRP; ADC98216) and prolactin  
CC -inducible protein (PIP; ADC98218), and their post-translationally  
CC modified forms. These proteins are secreted by salivary or sweat glands  
CC and bind to IgE autoantibodies, thereby activating mast cells and  
CC basophils. The invention also relates to antigenic peptide fragments of  
CC PRP or PIP; antibodies which bind to PRP or PIP; methods for diagnosing  
CC atopic dermatitis or for determining whether an individual is at risk of  
CC developing atopic dermatitis by determining the presence of PRP- or PIP-  
CC specific antibodies or immune complexes, or by quantifying histamine  
CC release; and sensitisation remedies for atopic dermatitis containing PRP  
CC and/or PIP or their peptides as the active ingredient. PRP, PIP and their  
CC antibodies are useful in diagnosing atopic dermatitis, or for determining  
CC whether an individual is at risk of developing atopic dermatitis. They  
CC are also useful in developing sensitisation remedies for the treatment of  
CC atopic dermatitis. The present sequence represents the specifically  
CC claimed human salivary acidic proline-rich phosphoprotein (PRP).

XX SQ Sequence 166 AA;

Query Match 100.0%; Score 53; DB 7; Length 166;

Best Local Similarity 100.0%; Pred. NO. 1.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRFQGPPQ 9

DB 122 RGRFQGPPQ 130

RESULT 14

RAB48774  
ID AAB48774 standard; peptide; 8 AA.

XX AC AAB48774;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 106-113), SEQ ID NO:4.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.

XX OS Homo sapiens.

XX PN WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SE000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX PA (STRO/) STROEMBERG N.  
XX (JOHA/) JOHANSSON I.

XX PI Stroemberg N, Johansson I;

XX DR WPI; 2001-031923/04.

XX PT New oligopeptides comprising 2 arginine residues from degradation of

proline-rich proteins, useful for preventing dental caries.

Claim 4; Page 24; 36pp; English.

The invention relates to human PRP-1-derived oligopeptides (AAB48771-AAB48783) which contain at least two arginine residues and which protect against dental caries. PRPs (proline-rich proteins) are salivary proteins encoded by six clustered genes on chromosome 12p13.2 and are potential determinants of a person's susceptibility to dental caries. PRPs are degraded by Actinomyces and Streptococcus species to small peptide fragments. These are metabolised by oral bacteria for nutritional purposes, with certain bacterial species generating ammonia via the catabolism of arginine. The peptides of the invention, being arginine-rich, can also be converted to ammonia by these bacteria. The ammonia thus formed raises the pH at the dental surface, thereby protecting the teeth against caries. Sequences AAB48771-548783 represent the PRP-1-derived oligopeptides of the invention

Sequence 8 AA;

Query Match 90.6%; Score 48; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. NO. 1.4e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGGP 8  
| | | | |  
DB 1 RGRPQGGP 8

RESULT 15  
AAW03557  
ID AAW03557 standard; peptide; 44 AA.

AC AAW03557;

DT 10-DEC-1996 (first entry)

DE Human proline-rich saliva glycoprotein P-C.

DE Human; saliva; proline-rich glycoprotein; insulin; secretion; glucagon;  
KW perfusion pancreas model; Wistar rat; diabetes.

XX Homo sapiens.

XX JP08092281-A.

XX 09-APR-1996.

XX 27-SEP-1994; 94JP-00257540.

XX 27-SEP-1994; 94JP-00257540.

XX (TEIK ) TEIKOKU HORMONE MFG CO LTD.

XX WPI; 1996-236094/24.

XX New octadeca-peptide for treating diabetes - acts as insulin secretion  
PT promoter and glucagon secretion inhibitor.

XX Disclosure; Page 2; 7pp; Japanese.

XX This is the sequence of the human proline-rich saliva glycoprotein P-C  
XX which was used to synthesise an octadecapeptide (AAW03556) which has an  
XX insulin secretion promoting activity and a glucagon secretion inhibiting  
XX activity. The peptide was generated by usual peptide synthesis method and  
XX opt. contains Boc protective gps. at the N-terminal His and on the  
XX residue during synthesis. These are removed by cleavage with HCl for 10  
XX mins on ice. The insulin secretion promoting activity and glucagon  
XX secretion inhibiting activity were measured in a perfusion pancreas model  
XX in a male Wistar rat. Insulin secretion was 1.78 times than normal and  
XX glucagon secretion was 0.77 times normal levels. The peptide can thus be  
XX used to treat diabetes

SQ Sequence 44 AA;

Query Match 90.6%; Score 48; DB 2; Length 44;  
Best Local Similarity 100.0%; Pred. NO. 2.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGGPP 9  
| | | | |  
DB 1 GRPQGGPP 8

Search completed: April 6, 2004, 16:06:38  
Job time : 51.3832 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 35.3271 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-5  
Perfect score: 53  
Sequence: 1 RGRPGPPQ 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	132	9	US-09-864-761-43644
2	53	100.0	166	14	Sequence 43644, A
3	46	86.8	149	12	US-10-157-031-80
4	43	81.1	803	14	Sequence 80, Appl
5	41	77.4	130	12	US-10-424-599-171613
6	41	77.4	159	12	Sequence 12588, A
7	41	77.4	224	12	US-10-156-761-12958
8	40	75.5	92	14	Sequence 215801, A
9	40	75.5	274	9	US-10-424-599-215801
10	40	75.5	458	15	Sequence 219996, A
11	40	75.5	19723	15	Sequence 62999, A
12	39	73.6	150	12	US-10-425-114-62999
13	39	73.6	209	15	Sequence 29755, A
14	39	73.6	451	15	US-09-850-887-4
15	39	73.6	842	12	Sequence 4, Appl

16	38	71.7	37	12	US-10-424-599-227554
17	38	71.7	52	12	Sequence 277859, A
18	38	71.7	76	12	US-10-424-599-277859
19	38	71.7	110	12	Sequence 204459, A
20	38	71.7	116	12	US-10-424-599-193123
21	38	71.7	121	12	US-10-424-599-280434
22	38	71.7	178	14	Sequence 280434, A
23	38	71.7	178	14	US-10-424-599-271545
24	38	71.7	4952	15	Sequence 2, Appl
25	38	71.7	5008	15	Sequence 6, Appl
26	38	71.7	5065	14	US-10-029-137-6
27	38	71.7	5159	15	US-10-051-874-56
28	38	71.7	5262	15	US-10-085-198-112
29	38	71.7	5262	15	Sequence 112, Appl
30	37.5	70.8	292	12	US-10-051-874-165
31	37.5	70.8	345	12	Sequence 167, Appl
32	37.5	70.8	440	12	US-10-051-874-167
33	37	69.8	17	9	US-10-425-114-66317
34	37	69.8	24	9	US-10-425-114-39512
35	37	69.8	28	14	US-09-864-761-34639
36	37	69.8	41	14	US-09-802-674-7
37	37	69.8	45	9	Sequence 7, Appl
38	37	69.8	45	9	Sequence 3, Appl
39	37	69.8	53	12	Sequence 28071, A
40	37	69.8	78	12	Sequence 30, Appl
41	37	69.8	85	12	Sequence 249209, A
42	37	69.8	87	12	US-10-424-599-171786
43	37	69.8	114	10	Sequence 171786, A
44	37	69.8	114	10	US-10-425-114-51102
45	37	69.8	114	10	Sequence 181318, A

## ALIGNMENTS

## RESULT 1

US-09-864-761-43644  
; Sequence 43644, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30

RESULT 3  
US-10-424-599-171613  
; Sequence 171613, Application US/10424599

```

; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62999
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3115-038-D1_FLI.pep
US-10-425-114-62999

Query Match 77.4%; Score 41; DB 12; Length 224;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGRPGPPQ 9
Db 155 RGAPAGPPQ 163

RESULT 8
US-10-029-386-29755
; Sequence 29755, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29755
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR17.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.71
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.99
; OTHER INFORMATION: SWISSPROT HIT: O35569, EVALU 1.10e-02
US-10-029-386-23755

Query Match 75.5%; Score 40; DB 14; Length 92;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPGPPQ 9
Db 47 GRPGPPK 54

RESULT 9
US-09-850-887-4
; Sequence 4, Application US/09850887
; Patent No. US20020009778A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina
; TITLE OF INVENTION: THYROID AND PITUITARY MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO

```

```

; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/850,887
; FILING DATE: 07-May-2001
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 09/087,678
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0535 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: g206712
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-850-887-4

```

```

Query Match 75.5%; Score 40; DB 9; Length 274;
Best Local Similarity 87.5%; Pred. No. 3.7e-02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 GRPQGPQ 9
DB 263 GRPQGPQ 270

```

```

RESULT 10
US-10-369-493-2375
; Sequence 2375, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2375
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(458)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-2375

```

```

Query Match 75.5%; Score 40; DB 15; Length 458;
Best Local Similarity 87.5%; Pred. No. 5.9e-02;

```

```

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GRPQGPQ 9
DB 395 GRPQGPQ 402

```

```

RESULT 11
US-10-084-846A-5
; Sequence 5, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 5
; LENGTH: 19723
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5

```

```

Query Match 75.5%; Score 40; DB 15; Length 19723;
Best Local Similarity 77.8%; Pred. No. 1.7e-04;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 RGRPQGPQ 9
DB 10188 RGRPQGPQ 10196

```

```

RESULT 12
US-10-424-599-146985
; Sequence 146985, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 146985
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_103748C.1.pep
US-10-424-599-146985

```

```

Query Match 73.6%; Score 39; DB 12; Length 150;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 RGRPQGPQ 9

```

Db 104 RGRPGGPPK 112

## RESULT 13

US-10-108-260A-4370  
; Sequence 4370, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20040005560A1e1 full length cdNA  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 4370  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-4370

Query Match 73.6%; Score 39; DB 15; Length 209;  
Best Local Similarity 87.5%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGGPP 8

Db 119 RGRPGGHP 126

## RESULT 14

US-10-266-829-132  
; Sequence 132, Application US/10266829  
; Publication No. US20030220489A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 29 Human secreted proteins  
; FILE REFERENCE: P2041P1  
; CURRENT APPLICATION NUMBER: US/10/266,829  
; CURRENT FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: 09/756,168  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: PCT/US00/19735  
; PRIOR FILING DATE: 2000-07-20  
; PRIOR APPLICATION NUMBER: 60/145,220  
; PRIOR FILING DATE: 1999-07-23  
; NUMBER OF SEQ ID NOS: 146  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 132  
; LENGTH: 451  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (118)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-266-829-132

Query Match 73.6%; Score 39; DB 15; Length 451;  
Best Local Similarity 77.8%; Pred. No. 8e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGRPGGPPQ 9

Db 154 RGRPGGPPQ 162

## RESULT 15

US-10-112-944-373  
; Sequence 373, Application US/10112944  
; Publication No. US20040048249A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom

; APPLICANT: Yang, Yonghong  
; APPLICANT: Weng, Gezhi  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Ghosh, Malabika  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wang, Zhiwei  
; TITLE OF INVENTION: No. US20040048249A1e1 Nucleic Acids and  
; TITLE OF INVENTION: Secreted Polypeptides  
; FILE REFERENCE: 805A  
; CURRENT APPLICATION NUMBER: US/10/112,944  
; CURRENT FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: US 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; PRIOR APPLICATION NUMBER: US 09/491,404  
; PRIOR FILING DATE: 2000-01-25  
; PRIOR APPLICATION NUMBER: US 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; PRIOR APPLICATION NUMBER: US 09/515,126  
; PRIOR FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: US 09/519,705  
; PRIOR FILING DATE: 2000-03-07  
; PRIOR APPLICATION NUMBER: US 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: US 09/552,929  
; PRIOR FILING DATE: 2000-04-18  
; PRIOR APPLICATION NUMBER: US 09/577,408  
; PRIOR FILING DATE: 2000-05-18  
; NUMBER OF SEQ ID NOS: 924  
; SOFTWARE: pt FL\_genes Version 5.0  
; SEQ ID NO 373  
; LENGTH: 842  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-112-944-373

Query Match 73.6%; Score 39; DB 12; Length 842;  
Best Local Similarity 75.0%; Pred. No. 1.4e+03;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPGGPPQ 9

Db 456 GRPGGPPR 463

Search completed: April 6, 2004, 17:05:53  
Job time : 36.3271 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 11.215 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-6

Perfect score: 58

Sequence: 1 RGRPQGPQQ 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	100.0	166	1	PIHUSC
2	58	100.0	166	2	salivary proline-r
3	58	100.0	171	2	salivary proline-r
4	50	86.2	128	2	proline-rich phosph
5	50	86.2	188	2	basic proline-rich
6	50	86.2	251	1	salivary proline-r
7	50	86.2	310	1	salivary proline-r
8	50	86.2	392	1	salivary proline-r
9	48	82.8	295	2	salivary proline-r
10	47	81.0	797	2	proline-rich prote
11	44	75.9	76	2	hypothetical prote
12	44	75.9	117	2	basic proline-rich
13	44	75.9	170	2	proline-rich prote
14	44	75.9	206	1	proline-rich prote
15	44	75.9	212	2	acidic proline-ric
16	44	75.9	212	2	proline-rich prote
17	43	74.1	188	2	proline-rich prote
18	43	74.1	317	2	proline-rich prote
19	43	74.1	646	2	hypothetical prote
20	43	74.1	1089	2	serine-repeat anti
21	42	72.4	231	2	hypothetical prote
22	41	70.7	125	2	synaptobrevin - lo
23	41	70.7	220	2	proline-rich prote
24	41	70.7	1276	2	SREBP cleavage act
25	40	69.0	147	2	proline-rich prote
26	40	69.0	164	2	proline-rich prote
27	40	69.0	172	2	proline-rich prote
28	40	69.0	183	2	acidic proline-ric
29	40	69.0	204	2	proline-rich prote

## ALIGNMENTS

### RESULT 1

PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N/Alternate names: salivary acidic proline-rich protein PRH2

N/Contains: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotein

C/Species: Homo sapiens (man)

C/Date: 31-Mar-1981 #sequence revision 12-Apr-1996 #text change 08-Dec-2000

C/Accession: A25372; A19803; B57868; A92277; A92254; A94425; A91954; S02564; S02563; JP

R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A/Title: Differential RNA splicing and post-translational cleavages in the human saliva

A/Reference number: A92492; MUID:85289325; PMID:2993301

A/Accession: A25372

A/Molecule type: mRNA

A/Residues: 1-166 <MAE>

A/Cross-references: GB:K03202; NID:G190461; PIDN:AAA60183.1; PID:G190482

R/Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A/Title: Primary structure of the active tryptic fragments of human and monkey saliva

A/Reference number: A91757; MUID:81191179; PMID:7228490

A/Accession: A19803

A/Molecule type: protein

A/Residues: 17-46 <SCH>

R/Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A/Title: Structures of two HasIII-type genes in the human salivary proline-rich protein

A/Reference number: A57868; MUID:86196106; PMID:3009472

A/Accession: B57868

A/Molecule type: DNA

A/Residues: 1-166 <KIM>

A/Cross-references: GB:M13058; NID:G190513; PIDN:AAA98808.1; PID:G190514

R/Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5949, 1980

A/Title: The primary structure of a salivary calcium-binding proline-rich phosphoprotei

A/Reference number: A92277; MUID:80204368; PMID:7380845

A/Accession: A92277

A/Molecule type: protein

A/Residues: 17-19, N', 21-166 <WON>

A/Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite form

R/Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A/Title: The complete primary structure of a proline-rich phosphoprotein from human sal

A/Reference number: A92254; MUID:79173237; PMID:438215

A/Contents: protein A

A/Accession: A92254

A/Molecule type: protein

A/Residues: 17-19, N', 21-122 <WO2>

R/Schlesinger, D.H.; Hay, D.I.

In Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gross

A/Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A/Reference number: A94425

salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human

C/Species: Homo sapiens (man)  
 C/Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text change 20-Aug-1999  
 C/Accession: B25372; A57868; S02562; G38355; S06153; B27307  
 R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 11123-11130, 1985  
 A/Title: Differential RNA splicing and post-translational cleavages in the human saliva  
 A/Reference number: A92492; MUID:85289325; PMID:2993301  
 A/Accession: B25372  
 A/Molecule type: mRNA  
 A/Residues: 1-166 <MAE>  
 A/Cross-references: GB:K03203; NID:G190483; PIDN:AAA60184.1; PID:G190484  
 R/Kim, H.S.; Maeda, N.  
 J. Biol. Chem. 261, 6712-6718, 1986  
 A/Title: Structures of two HaellI-type genes in the human salivary proline-rich protein  
 A/Reference number: A57868; MUID:86196106; PMID:3009472  
 A/Accession: A57868  
 A/Molecule type: DNA  
 A/Residues: 1-166 <KIM>  
 A/Cross-references: GB:M13057; NID:G190511; PIDN:AAA98807.1; PID:G190512  
 R/Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluc  
 Biochem. J. 255, 15-21, 1988  
 A/Title: The primary structures of six human salivary acidic proline-rich proteins (PRP  
 A/Reference number: S02562; MUID:89061650; PMID:3196309  
 A/Accession: S02562  
 A/Molecule type: protein  
 A/Residues: 47-71 <HAY>  
 R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the co  
 A/Reference number: A38355; MUID:91190884; PMID:1849422  
 A/Accession: G38355  
 A/Molecule type: protein  
 A/Residues: 123-166 <KAU>  
 R/Robinson, R.; Kaufman, D.L.; Waye, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A/Title: Primary structure and possible origin of the non-glycosylated basic proline-ri  
 A/Reference number: S06153; MUID:90089384; PMID:2688632  
 A/Accession: S06153  
 A/Molecule type: protein  
 A/Residues: 123-166 <ROB>  
 R/Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A/Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich pr  
 A/Reference number: A27307; MUID:88074309; PMID:3687941  
 A/Contents: allele Pa  
 A/Accession: B27307  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: DNA  
 A/Residues: 17-41, 'L', 43-118, 'C', 120-166 <AZE>  
 A/Cross-references: EMBL:K03203  
 C/Genetics:  
 A/Gene: GDB:PRH1  
 A/Cross-references: GDB:119515; OMIM:168730  
 A/Map position: 12p13.2-12p13.2  
 A/Introns: 22/1; 34/1  
 A/17-122/Product: protein A #status experimental <PA>  
 F/17-46/Region: apatitic mineral binding  
 F/47-71/Product: PRP-3 #status experimental <PRP3>  
 F/123-166/Product: peptide P-C #status experimental <PPC>  
 F/17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime  
 F/24,38/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 100.0%; Score 58; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.045;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGPQQ 10  
 |||||  
 DB 122 RGRPQGPQQ 131

RESULT 3  
 A27307  
 proline-rich phosphoprotein (gene PRH1, Db allele) - human  
 N/Alternate names: salivary acidic proline-rich protein  
 C/Species: Homo sapiens (man)

A/Accession: A94425  
 A/Molecule type: protein  
 A/Residues: 17-122 <SC2>  
 A/Note: the authors call this protein PRP-4  
 R/Isemura, S.; Saitoh, E.; Sanada, K.  
 J. Biochem. 87, 1071-1077, 1980  
 A/Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relat  
 A/Reference number: A91954; MUID:80227634; PMID:7390979  
 A/Contents: peptide P-C  
 A/Accession: A91954  
 A/Molecule type: protein  
 A/Residues: 123-166 <IGE>  
 R/Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluc  
 Biochem. J. 255, 15-21, 1988  
 A/Title: The primary structures of six human salivary acidic proline-rich proteins (PRP  
 A/Reference number: S02562; MUID:89061650; PMID:3196309  
 A/Accession: S02564  
 A/Molecule type: protein  
 A/Residues: 17-166 <HAY>  
 A/Accession: S02563  
 A/Molecule type: protein  
 A/Residues: 47-71 <HA2>  
 R/Schlesinger, D.H.; Hay, D.I.  
 Int. J. Pept. Protein Res. 27, 373-379, 1986  
 A/Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibit  
 A/Reference number: JP0106; MUID:86222916; PMID:3710693  
 A/Accession: JP0106  
 A/Molecule type: protein  
 A/Residues: 17-161, 'Q', 163-166 <SC3>  
 A/Experimental source: parotid gland  
 R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov  
 A/Reference number: A38355; MUID:91190884; PMID:1849422  
 A/Accession: G38355  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 123-166 <KAU>  
 R/Robinson, R.; Kaufman, D.L.; Waye, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A/Title: Primary structure and possible origin of the non-glycosylated basic proline-ri  
 A/Reference number: S06153; MUID:90089384; PMID:2688632  
 A/Accession: S06153  
 A/Molecule type: protein  
 A/Residues: 123-166 <ROB>  
 C/Comment: The proposed biological functions are a highly potent inhibitor of crystal gr  
 C/Genetics:  
 A/Gene: GDB:PRH2  
 A/Cross-references: GDB:119516; OMIM:168790  
 A/Map position: 12p13.2-12p13.2  
 A/Introns: 22/1; 34/1  
 C/Suprafamily: proline-rich protein  
 C/Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva  
 F/1-16/Domain: signal sequence #status predicted <SIG>  
 F/17-166/Product: protein C #status experimental <PRC>  
 F/17-122/Product: protein A #status experimental <PA>  
 F/17-46/Region: apatitic mineral binding  
 F/47-71/Product: PRP-3 #status experimental <PRP3>  
 F/123-166/Product: peptide P-C #status experimental <PPC>  
 F/17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime  
 F/24,38/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 100.0%; Score 58; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.045;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGPQQ 10  
 |||||  
 DB 122 RGRPQGPQQ 131

RESULT 2  
 B25372

C>Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 29-Aug-1997  
C/Accession: A27307  
R/Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
Am. J. Hum. Genet. 41, 1035-1047, 1987  
A/Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
A/Reference number: A27307; MUID:88074309; PMID:3687941  
A/Accession: A27307  
A/Status: nucleic acid sequence not shown  
A/Molecule type: DNA  
A/Residues: 1-171 <AZE>  
A/Cross-references: EMBL:K03203  
C/Genetics:  
A/Genes: GDB:PRH1  
A/Cross-references: GDB:119515; OMIM:168730  
A/Map position: 12p13.2-12p13.2  
C/Superfamily: proline-rich protein  
C/Keywords: phosphoprotein

Query Match 100.0%; Score 58; DB 2; Length 171;  
Best Local Similarity 100.0%; Pred. No. 0.046;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRPQGPQQ 10  
DB 127 RRPQGPQQ 136

RESULT 4  
D38355  
basic proline-rich peptide IB-8a - human (fragments)  
C/Species: Homo sapiens (man)  
C/Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 12-Apr-1995  
C/Accession: D38355  
R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
Biochemistry 30, 3351-3356, 1991  
A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov  
A/Reference number: A38355; MUID:91190884; PMID:1849422  
A/Accession: D38355  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 1-128 <KAU>  
C/Superfamily: proline-rich protein

Query Match 86.2%; Score 50; DB 2; Length 128;  
Best Local Similarity 88.9%; Pred. No. 0.6;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10  
DB 99 GRPQGPQQ 107

RESULT 5  
JH0481  
basic proline-rich protein MnP4 - crab-eating macaque  
C/Species: Macaca fascicularis (crab-eating macaque)  
C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Aug-1999  
R/Kin, H.H.; Kousvelari, E.E.; Ann, D.K.  
Gene 104, 219-226, 1991  
A/Title: Sequence and expression of the MnP4 gene encoding basic proline-rich protein in  
A/Reference number: JH0481; MUID:92009216; PMID:1916292  
A/Accession: JH0481  
A/Molecule type: DNA  
A/Residues: 1-188 <LIN>  
A/Cross-references: GB:M81322; GB:M61736; NID:G342284; PIDN:AAA36905.1; PID:G342285  
A/Experimental source: salivary gland  
A/Accession: JH0482  
A/Molecule type: mRNA  
A/Residues: 1-188 <LIN>  
A/Cross-references: GB:M81321; GB:M61735; NID:G342282; PIDN:AAA36904.1; PID:G342283  
C/Comment: This protein is a major constituent of both parotid gland and submandibular g  
C/Genetics:

A/Introns: 22/1; 34/1; 187/2  
C/Superfamily: proline-rich protein

Query Match 86.2%; Score 50; DB 2; Length 188;  
Best Local Similarity 88.9%; Pred. No. 0.86;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10  
DB 95 GRPQGPQQ 103

RESULT 6  
P1HUPP  
salivary proline-rich glycoprotein precursor PRB2 (validated) - human (fragment)  
N/Alternate names: basic proline-rich peptide IB-8c precursor; proline-rich protein (cl  
N/Contains: basic proline-rich peptide IB-4; basic proline-rich peptide P-F  
C/Species: Homo sapiens (man)  
C/Date: 15-Nov-1984 #sequence\_revision 12-Apr-1996 #text\_change 08-Dec-2000  
R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
J. Biol. Chem. 260, 11123-11130, 1985  
A/Title: Differential RNA splicing and post-translational cleavages in the human saliva  
A/Reference number: A92492; MUID:85289325; PMID:2993301  
A/Accession: E25372  
A/Molecule type: mRNA  
A/Residues: 1-251 <MAE>  
A/Cross-references: GB:K03208; NID:G190509; PIDN:AAA60189.1; PID:G190510  
R/Mamula, P.W.; Morley, D.J.; Larsen, S.H.; Karn, R.C.  
Biochem. Genet. 26, 165-175, 1988  
A/Title: Expression of human salivary protein genes.  
A/Reference number: A60827; MUID:88240287; PMID:3288192  
A/Accession: A60827  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 175-251 <MAM>  
R/Saitoh, E.; Isemura, S.; Sanada, K.  
J. Biochem. 93, 883-888, 1983  
A/Title: Complete amino acid sequence of a basic proline-rich peptide, P-F, from human  
A/Reference number: A03294; MUID:83265674; PMID:6874669  
A/Accession: A03294  
A/Molecule type: protein  
A/Residues: 134-194 <SAI>  
A/Experimental source: saliva  
R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
Biochemistry 30, 3351-3356, 1991  
A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the co  
A/Reference number: A38355; MUID:91190884; PMID:1849422  
A/Accession: B38355  
A/Molecule type: protein  
A/Residues: 134-194 <KAU>  
A/Experimental source: saliva  
A/Note: this peptide, which is closely related to that of peptide P-E, contains three 2  
A/Accession: A38355  
A/Molecule type: protein  
A/Residues: 10-67, R' <KA2>  
A/Accession: F38355  
A/Molecule type: protein  
A/Residues: 196-251 <KA3>  
C/Genetics:  
A/Genes: GDB:PRB2  
A/Cross-references: GDB:119512; OMIM:168810  
A/Map position: 12p13.2-12p13.2  
C/Superfamily: proline-rich protein  
C/Keywords: glycoprotein; saliva; tandem repeat  
F:134-194/Product: basic proline-rich peptide IB-4 #status experimental <MAT1>  
F:196-251/Product: basic proline-rich peptide IB-4 #status experimental <MAT2>  
F:3, 65, 107/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.2%; Score 50; DB 1; Length 251;  
Best Local Similarity 88.9%; Pred. No. 1.1;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;



```

QY      2 GRPQPPQQ 10
Db      199 GRPQPPQQ 207

RESULT 7
PIHUSD
salivary proline-rich glycoprotein precursor PRB4 (large allele) [validated] - human
N:Contains: basic proline-rich protein IB-5; proline-rich peptide P-D
C:Species: Homo sapiens (man)
C>Date: 19-Feb-1984 #sequence revision 12-Apr-1996 #text change 08-Dec-2000
C:Accession: S03176; S10890; D25372; E38355; A03295; A61294; S62891
R:Lyons, K.M.; Stein, J.H.; Smithies, O.
Genetics 120, 267-278, 1988
A:Title: Length polymorphisms in human proline-rich protein genes generated by intragenic
A:Reference number: S02127; MUID:89121440; PMID:2851479
A:Accession: S03176
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 35-310 <LY1>
A:Cross-references: EMBL:X07715
A>Note: large allele
A:Accession: S03175
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 35-36, 'E', 38-112, 155-310 <LY2>
A:Cross-references: EMBL:X07704
A>Note: medium allele
A:Accession: S10890
A>Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 1-38, 60-112, 'T', 114-115, 'P', 117-121, 185-271, 'A', 273-310 <LY3>
A:Cross-references: EMBL:X07882; NID:G35647; PIDN:CAA30729.1; PID:G296670
R:Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.
J. Biol. Chem. 260, 11223-11230, 1985
A:Title: Differential RNA splicing and post-translational cleavages in the human salivary
A:Reference number: A92492; MUID:85289325; PMID:2993301
A:Accession: D25372
A:Molecule type: mRNA
R:Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
Biochemistry 30, 3351-3356, 1991
A:Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov
A:Reference number: A38355; MUID:91190884; PMID:1849422
A:Accession: E38355
A:Molecule type: protein
A:Residues: 241-254, 'KN', 257-310 <KAU>
R:Saich, E.; Isemura, S.; Sanada, K.
J. Biochem. 93, 495-502, 1983
A:Title: Complete amino acid sequence of a basic proline-rich peptide, P-D, from human P
A:Reference number: A03295; MUID:83186122; PMID:6841349
A:Accession: A03295
A:Molecule type: protein
A:Residues: 241-310 <SAI>
R:Shimomura, H.; Kanai, Y.; Sanada, K.
J. Biochem. 93, 857-863, 1983
A:Title: Amino acid sequences of glycopeptides obtained from basic proline-rich glycopro
A:Reference number: A61294; MUID:83265671; PMID:6874667
A:Accession: A61294
A:Molecule type: protein
A:Residues: 54-57, 'E', 59-73, 'R', 82-101 <SHI>
R:Charlton, A.J.; Baxter, N.J.; Lilley, T.H.; Haslam, E.; McDonald, C.J.; Williamson, M.
FEBS Lett. 382, 289-292, 1996
A:Title: Tannin interactions with a full-length human salivary proline-rich protein disp
A:Reference number: S62891; MUID:96184506; PMID:8605987
A:Accession: S62891
A:Molecule type: protein
A:Residues: 241-252 <CRA>
A>Note: amino end of peptide designated basic proline-rich protein IB-5
A>Note: it is unclear from the peptide sequence whether this is a product of the PRB2 (P
C:Genetics:
A:Gene: GDB:PRB4
A:Cross-references: GDB:119514; OMIM:180990

```

---

```

A:Map position: 12p13.2-12p13.2
A:introns: 22/1; 34/1
A>Note: the list of introns may be incomplete
C:Superfamily: proline-rich protein
C:Keywords: glycoprotein; saliva; tandem repeat
F:1-16/Domain: signal sequence #status predicted <SIG>
F:241-310/Product: proline-rich peptide P-D #status experimental <MAT>
F:66,97,171/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:108,150,192,213,234/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      86.2%; Score 50; DB 1; Length 310;
Best Local Similarity 88.9%; Pred. No. 1-4;
Matches 8; Conservative 1; Mismatches 0; Gaps 0;

QY      2 GRPQPPQQ 10
Db      244 GRPQPPQQ 252

RESULT 8
PIHUS6
salivary proline-rich phosphoprotein precursor PRB1 (large allele) [validated] - human
N:Contains: peptide IB-1; peptide P-E (peptide IB-9); peptide P-F; peptide P-H
C:Species: Homo sapiens (man)
C>Date: 04-Dec-1986 #sequence revision 12-Apr-1996 #text change 08-Dec-2000
C:Accession: B40750; C40750; A40750; C25372; S02128; A03293; A90502; A91974; A0
R:Azen, E.A.; Latreille, P.; Niece, R.L.
Am. J. Hum. Genet. 53, 264-278, 1993
A:Title: PRB1 gene variants coding for length and null polymorphisms among human saliva
A:Reference number: A40750; MUID:93304421; PMID:8317492
A:Accession: B40750
A:Molecule type: DNA
A:Residues: 35-392 <AZE>
A:Cross-references: GB:S62941
A:Experimental source: subject C.J. (large allele)
A:Accession: C40750
A:Molecule type: DNA
A:Residues: 35-127, 'R', 129-148, 'R', 150-151, 153-187, 'K', 189-272, 'S', 274-336, 'S', 338-392
A:Cross-references: GB:S62929
A:Experimental source: subject M.V.O. (large allele)
A:Accession: A40750
A:Molecule type: DNA
A:Residues: 35-183, 245-270, 'Q', 272-392 <AZ3>
A:Cross-references: GB:S62928
A:Experimental source: subject C.J. (medium allele)
A>Note: authors translated the codon CAA for residue 272 as Arg
R:Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.
J. Biol. Chem. 260, 11123-11130, 1985
A:Title: Differential RNA splicing and post-translational cleavages in the human saliva
A:Reference number: A92492; MUID:85289325; PMID:2993301
A:Accession: C25372
A:Molecule type: mRNA
A:Residues: 1-183, 245-392 <MAE>
A:Cross-references: GB:K03204; NID:G190485; PIDN:AAA60185.1; PID:G190486
A>Note: alternatively splice forms lacking portions of the repeat region were also found
R:Lyons, K.M.; Stein, J.H.; Smithies, O.
Genetics 120, 267-278, 1988
A:Title: Length polymorphisms in human proline-rich protein genes generated by intragen
A:Reference number: S02127; MUID:89121440; PMID:2851479
A:Accession: S02128
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 35-127, 250-273, 'R', 275-277, 'R', 279-336, 'S', 338-392 <LYO>
A:Cross-references: EMBL:X07517
A:Accession: S02127
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 35-183, 245-392 <LY2>
A:Cross-references: EMBL:X07516
R:Kauffman, D.; Hofmann, T.; Bennick, A.; Keller, P.
Biochemistry 25, 2387-2392, 1986
A:Title: Basic proline-rich proteins from human parotid saliva: complete covalent struc
A:Reference number: A90502; MUID:86243355; PMID:3521730

```

A;Accession: A03293  
 A;Molecule type: protein  
 A;Residues: 17-38,'AP',41-51,92-148,'R',150-152 <KA2>  
 A;Note: among nine basic proline-rich peptides isolated from the saliva, this peptide is  
 A;Accession: A90502  
 A;Molecule type: protein  
 A;Residues: 275-336,'S',338-392 <KAU>  
 R;Saitoh, E.; Isemura, S.; Sanada, K.  
 J. Biochem. 94, 1991-1999, 1993  
 A;Title: Further fractionation of basic proline-rich peptides from human parotid saliva  
 A;Reference number: A91974; MUID:84161824; PMID:6671974  
 A;Contents: P-H  
 A;Accession: A91974  
 A;Molecule type: protein  
 A;Residues: 35-39,'P',41-84,'G',86,'R',87-154,'R',218-246,300-306,'T',308-329,'C',331-38  
 A;Accession: A05262  
 A;Molecule type: DNA  
 A;Residues: 'N',57-59,'A',61-69,334-336,'S',338-339,'R',341-392 <AZ5>  
 R;Kauffman, D.; Wong, R.; Bennick, A.; Keller, P.  
 Biochemistry 21, 6558-6562, 1982  
 A;Title: Basic proline-rich proteins from human parotid saliva: complete covalent structure  
 A;Reference number: A90464; MUID:83101329; PMID:6924859  
 A;Contents: IB-9  
 A;Accession: A90464  
 A;Molecule type: protein  
 A;Residues: 92-127,'R',129-148,'R',150-152 <KA3>  
 R;Isemura, S.; Saitoh, E.; Sanada, K.  
 J. Biochem. 91, 2067-2075, 1982  
 A;Title: Fractionation and characterization of basic proline-rich peptides of human parotid  
 A;Contents: P-E  
 A;Reference number: A91966; MUID:83007119; PMID:7118863  
 A;Accession: A91966  
 A;Molecule type: protein  
 A;Residues: 92-127,'R',129-148,'R',150-152 <ISE>  
 C;Comment: This peptide contains 21-residue repeats, two of which have internal 7-residue  
 C;Genetics:  
 A;Gene: GDB:PRB1  
 A;Cross-references: GDB:119511; OMIM:180989  
 A;Map position: 12p13.2-12p13.2  
 A;Note: each of the tandem repeats contains a candidate splice acceptor site, and several  
 C;Superfamily: proline-rich protein  
 C;Keywords: alternative splicing; duplication; parotid gland; phosphoprotein; pyroglutamate  
 F;1-16/Domain: signal sequence #status predicted <SIG>  
 F;17-51,92-152/Product: basic proline-rich peptide IB-1 #status experimental <IB1>  
 F;92-152/Product: basic proline-rich peptide P-E #status experimental <PPE>  
 F;275-392/Product: basic proline-rich peptide IB-6 #status experimental <PIB6>  
 F;275-335/Product: basic proline-rich peptide P-F #status experimental <PPF>  
 F;337-392/Product: basic proline-rich peptide P-H #status experimental <PPH>  
 F;17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experiment  
 F;24/Binding site: phosphates (Ser) (covalent) #status experimental  
 Query Match 86.2%; Score 50; DB 1; Length 392;  
 Best Local Similarity 88.9%; Pred. No. 1.7;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 GRPQGPQQ 10  
 Db 238 GRPQGPQQ 246  
 RESULT 9  
 B48013  
 Proline-rich proteoglycan 2 precursor, parotid - rat  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 02-Jun-1995 #sequence\_revision 02-Jun-1995 #text\_change 20-Aug-1999  
 C;Accession: B48013  
 R;Castle, A.M.; Castle, J.D.  
 J. Biol. Chem. 268, 20490-20496, 1993  
 A;Title: Novel secretory proline-rich proteoglycans from rat parotid. Cloning and charac  
 A;Reference number: A48013; MUID:93388626; PMID:8376404  
 A;Accession: B48013  
 A;Status: Preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-295 <CAS>  
 A;Cross-references: GB:L17318; NID:g310199; PIDN:AAA03074.1; PID:g310200  
 C;Superfamily: proline-rich protein  
 C;Keywords: extracellular protein; glycoprotein; tandem repeat  
 Query Match 82.8%; Score 48; DB 2; Length 295;  
 Best Local Similarity 88.9%; Pred. No. 2.6;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 GRPQGPQQ 10  
 Db 273 GNPQGPQQ 281  
 RESULT 10  
 S53590  
 Hypothetical protein YCR077c - yeast (Saccharomyces cerevisiae)  
 A;Alternate names: hypothetical protein YCR078c  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 19-May-1995 #sequence\_revision 01-Dec-1995 #text\_change 29-Oct-1999  
 C;Accession: S53590; S55867; S19491; S19492  
 R;Jimenez, A.  
 Submitted to the EMBL Data Library, December 1992  
 A;Reference number: S53589  
 A;Accession: S53590  
 A;Molecule type: DNA  
 A;Residues: 1-797 <JIM>  
 A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42266.1; PID:e264564; PID:g19072  
 R;Rodriguez-Cousino, N.; Lill, R.; Neupert, W.; Court, D.A.  
 Yeast 11, 581-585, 1995  
 A;Title: Identification and initial characterization of the cytosolic protein Ycr77p.  
 A;Reference number: S55867; MUID:95373282; PMID:7645349  
 A;Accession: S55867  
 A;Molecule type: DNA  
 A;Residues: 1-797 <ROD>  
 A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42266.1; PID:e264564; PID:g19072  
 A;Note: this is a correction to the sequence in reference S19486  
 R;Ballesta, J.P.G.; Franco, L.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E.  
 Submitted to the Protein Sequence Database, March 1992  
 A;Reference number: S19486  
 A;Accession: S19491  
 A;Molecule type: DNA  
 A;Residues: 135-641,'PH' <BAL>  
 A;Cross-references: EMBL:X59720; MIPS:YCR077c  
 A;Note: this sequence was incomplete due to a frameshift error  
 A;Note: this sequence has been revised in reference S55867  
 A;Accession: S19492  
 A;Molecule type: DNA  
 A;Residues: 'MNV',64,'SGRYLLTPDDLHP',79-80,'PSSKPOEQVYVPDKPWEH',83-132,'RKWP',137,'YSC  
 A;Cross-references: EMBL:X59720  
 A;Note: this sequence has been revised in reference S55867  
 A;Note: it was assumed that this is protein YCR078c  
 C;Genetics:  
 A;Gene: SGD:PAT1  
 A;Cross-references: MIPS:YCR077c; SGD:S0000673  
 A;Map position: 3R  
 Query Match 81.0%; Score 47; DB 2; Length 797;  
 Best Local Similarity 80.0%; Pred. No. 9.3;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GRPQGPQQ 10  
 Db 168 QGLPQGPQQ 177  
 RESULT 11

```

C38355
basic proline-rich peptide II-2 - human
C:Species: Homo sapiens (man)
C:Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 12-Apr-1995
C:Accession: C38355
R:Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
Biochemistry 30, 3351-3356, 1991
A:Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov
A:Reference number: A38355; MUID:91190884; PMID:1849422
A:Accession: C38355
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-76 <XAU>
C:Superfamily: proline-rich protein

Query Match 75.9%; Score 44; DB 2; Length 76;
Best Local Similarity 77.8%; Pred. No. 3.1;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 2 GRPQGPPQ 10
DB 40 GKPGPPPPQ 48

RESULT 12
D40750
proline-rich protein PRB1/28 (EA) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 03-May-1996
C:Accession: D40750
R:Azen, E.A.; Latreille, P.; Niece, R.L.
Am. J. Hum. Genet. 53, 264-278, 1993
A:Title: PRB1 gene variants coding for length and null polymorphisms among human salivar
A:Reference number: A40750; MUID:93304421; PMID:8317492
A:Accession: D40750
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-117 <AZE>
A:Cross-references: GB:S62930
C:Superfamily: proline-rich protein

Query Match 75.9%; Score 44; DB 2; Length 117;
Best Local Similarity 77.8%; Pred. No. 4.6;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 2 GRPQGPPQ 10
DB 21 GKPGPPPPQ 29

RESULT 13
A48013
proline-rich proteoglycan 1 precursor, parotid - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 20-Aug-1999
C:Accession: A48013
R:Castle, A.M.; Castle, J.D.
J. Biol. Chem. 268, 20490-20496, 1993
A:Title: Novel secretory proline-rich proteoglycans from rat parotid. Cloning and charac
A:Reference number: A48013; MUID:9338626; PMID:8376404
A:Accession: A48013
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-170 <CAS>
A:Cross-references: GB:L17317; NID:G310197; PIDN:AAA03073.1; PID:G310198
C:Superfamily: proline-rich protein
C:Keywords: extracellular protein; glycoprotein; tandem repeat

Query Match 75.9%; Score 44; DB 2; Length 170;
Best Local Similarity 77.8%; Pred. No. 6.6;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 2 GRPQGPPQ 10

```

```

Db 111 GKPGPPPPQ 119

RESULT 14
F1RT3
acidic proline-rich protein precursor - rat
N:Alternate names: PRP
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 22-Jun-1999
C:Accession: A03296
R:Ziemer, M.A.; Swain, W.F.; Rutter, W.J.; Clements, S.; Ann, D.K.; Carlson, D.M.
J. Biol. Chem. 259, 10475-10480, 1984
A:Title: Nucleotide sequence analysis of a proline-rich protein cDNA and peptide homolo
A:Reference number: A03296; MUID:84289443; PMID:6547951
A:Accession: A03296
A:Molecule type: mRNA
A:Residues: 1-206 <ZIE>
A:Cross-references: GB:X02247; NID:G206395; PIDN:AAA41949.1; PID:G206396
C:Comment: This protein contains six 18- to 19-residue repeats.
C:Comment: This protein may protect teeth by binding to tannins.
C:Superfamily: proline-rich protein
C:Keywords: duplication; parotid gland; saliva; tandem repeat
F:1-13/Domain: signal sequence #status predicted <SIG>
F:14-206/Product: acidic proline-rich protein #status predicted <MAT>
F:80-189/Region: 18-residue repeats

Query Match 75.9%; Score 44; DB 1; Length 206;
Best Local Similarity 77.8%; Pred. No. 7.8;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 2 GRPQGPPQ 10
DB 148 GKPGPPPPQ 156

RESULT 15
B36298
proline-rich protein PRB3s (cys) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 29-Aug-1997
C:Accession: B36298
R:Azen, E.A.; Minaguchi, K.; Latreille, P.; Kim, H.S.
Am. J. Hum. Genet. 47, 686-697, 1990
A:Title: Alleles at the PRB3 locus coding for a disulfide-bonded human salivary proline
A:Reference number: A36298; MUID:91022705; PMID:2171329
A:Accession: B36298
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-212 <AZE>
C:Genetics:
A:Gene: GDB:PRB3
A:Cross-references: GDB:119513; OMIM:168840
A:Map position: 12p13.2-12p13.2
C:Superfamily: proline-rich protein

Query Match 75.9%; Score 44; DB 2; Length 212;
Best Local Similarity 77.8%; Pred. No. 8;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 2 GRPQGPPQ 10
DB 147 GKPGPPPPQ 155

Search completed: April 6, 2004, 16:16:51
Job time : 11.215 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 6.4486 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-6

Perfect score: 58

Sequence: 1 RGRPGPPQQ 10

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	100.0	166	1 PRPC_HUMAN	P02810 homo sapien
2	50	86.2	234	1 PRPM_HUMAN	P10161 homo sapien
3	50	86.2	247	1 PRB4_HUMAN	P10163 homo sapien
4	50	86.2	251	1 PRP2_HUMAN	P02812 homo sapien
5	50	86.2	276	1 PRP1_HUMAN	P10162 homo sapien
6	50	86.2	331	1 PRP1_HUMAN	P04280 homo sapien
7	47	81.0	797	1 PAT1_YEAST	P25644 saccharomyc
8	44	75.9	61	1 PRP2_HUMAN	P02811 homo sapien
9	44	75.9	96	1 PRP5_HUMAN	P04281 homo sapien
10	44	75.9	174	1 PRPP_HUMAN	P81489 homo sapien
11	44	75.9	206	1 PRP3_RAT	P04474 rattus norv
12	42	72.4	124	1 CES4_HUMAN	Q9bqx7 homo sapien
13	41	70.7	125	1 SYB_LOLPE	P47194 loligo peal
14	41	70.7	1276	1 SCAP_CRIGR	P97260 cricetus
15	40	69.0	172	1 PRP2_RAT	P10164 rattus norv
16	40	69.0	183	1 PRPH_MESAU	P06680 mesocricetu
17	40	69.0	842	1 ORP7_HUMAN	Q9bfz2 homo sapien
18	40	69.0	1030	1 FBP1_DROME	Q04691 drosophila
19	40	69.0	2716	1 CSA_DROME	Q8in94 drosophila
20	39	67.2	367	1 CSP_PLAYAO	P06914 plasmodium
21	39	67.2	1138	1 BMK3_MOUSE	Q91296 mus musculu
22	39	67.2	1161	1 BMK2_HUMAN	Q9nsv1 homo sapien
23	39	67.2	2911	1 FBH2_HUMAN	P35556 homo sapien
24	38	65.5	11	1 BPP_AGGHP	P04562 agkistrodon
25	38	65.5	296	1 PMP3_MOUSE	P05143 mus musculu
26	38	65.5	416	1 P12R_RAT	P43253 rattus norv
27	38	65.5	543	1 PKAA_STRCO	P54739 streptomyce
28	38	65.5	551	1 BRF_MOUSE	P70459 mus musculu
29	38	65.5	635	1 PARF_CAUCR	O54479 caulobacter
30	38	65.5	5065	1 EPPL_HUMAN	P58107 homo sapien
31	38	65.5	5262	1 MLL2_HUMAN	O14686 homo sapien
32	37	63.8	113	1 VAM2_XENLA	P47193 xenopus lae
33	37	63.8	189	1 SSB_PSESM	Q889u1 pseudomonas

RESULT 1	PRPC_HUMAN	STANDARD;	PRT;	166 AA.
ID	PRPC_HUMAN			
AC	P02810;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains: Peptide P-C].			
DE	PRH1 AND PRH2.			
GN	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).			
RX	MEDLINE=86196106; PubMed=3009472;			
RA	Kim H.-S., Maeda N.;			
RT	"Structures of two HaeIII-type genes in the human salivary proline-rich protein multigene family.";			
RL	J. Biol. Chem. 261:6712-6718(1986).			
RN	[2]			
RN	SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).			
RX	MEDLINE=85289325; PubMed=2993301;			
RA	Maeda N., Kim H.-S., Azen E.A., Smithies O.;			
RT	"Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.";			
RL	J. Biol. Chem. 260:11123-11130(1985).			
RN	[3]			
RP	SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).			
RX	MEDLINE=89061650; PubMed=3196309;			
RA	Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K.,			
RA	Madapalliam G., Schluckebier S.K.;			
RT	"The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";			
RL	Biochem. J. 255:15-21(1988).			
RN	[4]			
RP	SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).			
RX	MEDLINE=88074309; PubMed=3687941;			
RA	Azen E.A., Kim H.-S., Goodman P., Flynn S., Maeda N.;			
RT	"Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich proteins Pa, Db, and Pif.";			
RL	Am. J. Hum. Genet. 41:1035-1047(1987).			
RN	[5]			
RP	SEQUENCE OF 17-166 (PRP-2).			
RX	MEDLINE=86222916; PubMed=3710693;			
RA	Schlesinger D.H., Hay D.I.;			
RT	"Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva.";			
RL	Int. J. Pept. Protein Res. 27:373-379(1986).			
RN	[6]			
RP	SEQUENCE OF 17-166 (PROTEIN C).			
RX	MEDLINE=80204368; PubMed=7380845;			
RA	Wong R.S.C., Bennick A.;			

Q02067 mus musculu  
P05142 mus musculu  
P56470 homo sapien  
Q9ubf1 homo sapien  
Q9q49 rhizobium m  
Q62151 mus musculu  
Q28173 bos taurus  
Q95466 homo sapien  
O18996 sus scrofa  
P27972 simian immu  
P21411 squirrel no  
Q63003 rattus norv

## ALIGNMENTS



```
SQ SEQUENCE 234 AA; 23676 MW; 310AFF13A44E747F CRC64;
Query Match 86.2%; Score 50; DB 1; Length 234;
Best Local Similarity 88.9%; Pred. No. 0.62;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPOGPPQQ 10
DB 168 GKPOGPPQQ 176

RESULT 3
PRB4_HUMAN STANDARD; PRT; 247 AA.
ID PRB4_HUMAN STANDARD; PRT; 247 AA.
AC P10163; P02813;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein precursor (Allele S).
GN PRB4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85289325; PubMed=2993301;
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT "Differential RNA splicing and post-translational cleavages in the
RT human salivary proline-rich protein gene system.";
RL J. Biol. Chem. 260:11123-11130(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89121440; PubMed=2851479;
RA Lyons K.M., Stein J.H., Smithies O.;
RT "Length polymorphisms in human proline-rich protein genes generated
RT by intragenic unequal crossing over.";
RL Genetics 120:267-278(1988).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03207; AAA60188.1; -
DR EMBL; X07882; CAA30729.1; -
DR PIR; S03176; PIHUSD.
DR Genew; HGNC:9340; PRB4.
DR MIM; 168730; -
DR MIM; 180990; -
KW Repeat; Parotid gland; Multigene family; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 247 SALIVARY PROLINE-RICH PROTEIN PO.
FT CHAIN 17 39 PROTEIN N1.
FT CHAIN 40 177 GLYCOSYLATED PROTEIN A.
FT CHAIN 37 57 MISSING (IN REF. 2).
FT CONFLICT 218 218 D -> A (IN REF. 2).
FT CONFLICT 218 218 D -> A (IN REF. 2).
SQ SEQUENCE 247 AA; 25108 MW; 6A1943E435161691 CRC64;

Query Match 86.2%; Score 50; DB 1; Length 247;
Best Local Similarity 88.9%; Pred. No. 0.65;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPOGPPQQ 10
DB 181 GKPOGPPQQ 189

RESULT 4
PRP2_HUMAN STANDARD; PRT; 251 AA.
ID PRP2_HUMAN STANDARD; PRT; 251 AA.
AC P02812; 1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 05, Last sequence update)
DT 13-AUG-1987 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein precursor (Clone CP7) [Contains: Basic
DE peptide P-F] (Fragment).
GN PRB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85289325; PubMed=2993301;
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT "Differential RNA splicing and post-translational cleavages in the
RT human salivary proline-rich protein gene system.";
RL J. Biol. Chem. 260:11123-11130(1985).
RN [2]
RP SEQUENCE OF 134-194.
RX MEDLINE=83265674; PubMed=6874669;
RA Saitoh E., Isemura S., Sanada K.;
RT "Complete amino acid sequence of a basic proline-rich peptide, P-F,
RT from human parotid saliva.";
RL J. Biochem. 93:883-888(1983).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03208; AAA60189.1; -
DR PIR; E25372; PIHUPF.
DR Genew; HGNC:9338; PRB2.
DR MIM; 168810; -
KW Repeat; Parotid gland; Multigene family.
FT NON_TER 1 1
FT CHAIN 134 194 BASIC PEPTIDE P-F.
FT REPEAT <1 9
FT REPEAT 10 71
FT REPEAT 72 133
FT REPEAT 134 195
FT REPEAT 134 195
SQ SEQUENCE 251 AA; 24641 MW; D779F590C0BFF30B CRC64;

Query Match 86.2%; Score 50; DB 1; Length 251;
Best Local Similarity 88.9%; Pred. No. 0.66;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPOGPPQQ 10
DB 199 GKPOGPPQQ 207

RESULT 5
PRPL_HUMAN STANDARD; PRT; 276 AA.
ID PRPL_HUMAN STANDARD; PRT; 276 AA.
AC P10162; P02813;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein PO (Allele K) [Contains: Peptide P-D]
DE (fragment).
GN PRB4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RX MEDLINE=89121439; PubMed=3220251;
RA Lyons K.M., Stein J.R., Smithies O.;
RT "Many protein products from a few loci: assignment of human salivary
RT proline-rich proteins to specific loci.";
RL Genetics 120:255-265(1989).
RN [2]
RP SEQUENCE OF 207-276.
RX MEDLINE=83186122; PubMed=6841349;
RA Saitoh E., Isemura S., Sanada K.;
RT "Complete amino acid sequence of a basic proline-rich peptide, P-D,
RT from human parotid saliva.";
RL J. Biochem. 93:495-502(1983).
RN [2]
RP SEQUENCE OF 207-276.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; X07715; CAA30543.1; ALT_SEQ.
DR PIR; S03176; PIHUSD.
DR MIM; 168730; -.
DR GO; GO:0005576; C:extracellular; NAS.
DR InterPro; IPR000637; AT_HOOk.
DR PRINTS; PR00929; AT_HOOk.
KW Repeat; Parotid gland; Multigene family.
FT NON_TER 1 1
FT CHAIN 207 276 PEPTIDE P-D.
SQ SEQUENCE 276 AA; 27816 MW; 9F494926C979441A CRC64;

Query Match 86.2%; Score 50; DB 1; Length 276;
Best Local Similarity 88.9%; Pred. No. 0.73;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGPQQ 10
Db 210 GKPQGPQQ 218

RESULT 6
PRP1 HUMAN
ID PRP1 HUMAN STANDARD; PRT; 331 AA.
AC P04280;
DT 20-MAR-1987 (Rel. 04, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein precursor (Clones CP3, CP4 and CP5)
DE [Contains: Basic peptide IB-6; Peptide P-H].
GN PRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85289325; PubMed=2993301;
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT "Differential RNA splicing and post-translational cleavages in the
RT human salivary proline-rich protein gene system.";
RL J. Biol. Chem. 260:11123-11130(1985).
RN [2]
RP SEQUENCE OF 214-331.
RX MEDLINE=86243355; PubMed=3521730;
RA Kauffman D., Hofmann T., Benrick A., Keller P.;
RT "Basic proline-rich proteins from human parotid saliva: complete
RT covalent structures of proteins IB-1 and IB-6.";
RL Biochemistry 25:2387-2392(1986).
RN [3]
RP SEQUENCE OF 276-331.
RX MEDLINE=84161824; PubMed=6671974;

```

```

RA Saitoh E., Isemura S., Sanada K.;
RT "Further fractionation of basic proline-rich peptides from human
RT parotid saliva and complete amino acid sequence of basic proline-rich
RT peptide P-H.";
RL J. Biochem. 94:1991-1997(1983).
RN [2]
RP SEQUENCE OF 207-276.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; K03204; AAA60185.1; -.
DR EMBL; K03205; AAA60186.1; -.
DR EMBL; K03206; AAA60187.1; -.
DR PIR; B40750; PIHUB6.
DR Genew; HGNC:9337; PRB1.
DR MIM; 180989; -.
KW Repeat; Parotid gland; Multigene family; Signal.
FT SIGNAL 1 16
FT CHAIN 214 331 PEPTIDE IB-6.
FT CHAIN 276 331 PEPTIDE P-H.
FT VARIANT 106 238 Missing (in clone CP-4).
FT VARIANT 106 258 /FTId=VAR_005561.
FT VARIANT 276 331 Missing (in clone CP-5).
FT CONFLICT 276 276 A -> S (IN REF. 2 AND 3).
SQ SEQUENCE 331 AA; 32596 MW; 3F481FF8BA39751 CRC64;

Query Match 86.2%; Score 50; DB 1; Length 331;
Best Local Similarity 88.9%; Pred. No. 0.87;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGPQQ 10
Db 177 GKPQGPQQ 185

RESULT 7
PAT1 YEAST
ID PAT1 YEAST STANDARD; PRT; 797 AA.
AC P25644;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Topoisomerase II-associated protein PAT1.
GN PAT1 OR YCR077C OR YCR77C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,
RA Sanz E.;
RN Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS AND CHARACTERIZATION.
RX MEDLINE=95373282; PubMed=7645349;
RA Rodriguez-Cousino N., Lill R., Neupert W., Court D.A.;
RT "Identification and initial characterization of the cytosolic protein
RT Ycr77p.";
RL Yeast 11:581-585(1995).
RN [3]
RP FUNCTION.
RX MEDLINE=97128274; PubMed=8972867;
RA Wang X., Watt P.M., Louis E.J., Borts R.H., Hickson I.D.;
RT "Pat1: a topoisomerase II-associated protein required for faithful
RT chromosome transmission in Saccharomyces cerevisiae.";
RL Nucleic Acids Res. 24:4791-4797(1996).
CC -!- FUNCTION: Necessary for accurate chromosome transmission during

```



cell division. Interacts with topoisomerase TOP2.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: TO S.POMBE SPBC1937.10C.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC ENBL; X59720; CAA42266.1; --  
 CC PIR; S53590; S53590.  
 CC Germonline; 138977; --  
 CC SGD; S0000673; PATL.  
 CC DR GO; GO:0000932; C:cytoplasmic mRNA processing body; IDA.  
 CC DR GO; GO:0005843; C:cytosolic small ribosomal subunit (sensu Eu. . .; IDA.  
 CC DR GO; GO:0007059; P:chromosome segregation; IMP.  
 CC DR GO; GO:0006446; P:regulation of translational initiation; IGI.  
 CC DOMAIN 195 200  
 CC SEQUENCE 797 AA; 88487 MW; 2A54F39AB3E75ECC CRC64;  
 CC  
 CC Query Match 81.0%; Score 47; DB 1; Length 797;  
 CC Best Local Similarity 80.0%; Pred. No. 5.9;  
 CC Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC QY 1 GRPQGPQQ 10  
 CC DB 168 QGLPQGPQQ 177  
 CC  
 CC RESULT 8  
 CC PRPP\_HUMAN STANDARD; PRT; 61 AA.  
 CC ID PRPP\_HUMAN  
 CC AC P02811;  
 CC DT 21-JUL-1986 (Rel. 01, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DE Basic proline-rich peptide P-E (IB-9).  
 CC OS Homo sapiens (Human)  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC OX NCBI\_TaxID=9606;  
 CC RN [1]  
 CC RP SEQUENCE  
 CC RC MEDLINE=83101329; PubMed=6924859;  
 CC RA Kaufman D., Wong R., Bannick A., Keller P.;  
 CC RT "Basic proline-rich proteins from human parotid saliva: complete  
 CC covalent structure of protein IB-9 and partial structure of protein  
 CC IB-6, members of a polymorphic pair.";  
 CC RL Biochemistry 21:6558-6562(1982).  
 CC RN [2]  
 CC RP SEQUENCE  
 CC RC TISSUE=Saliva;  
 CC RX MEDLINE=83007119; PubMed=7118863;  
 CC RA Isemura S., Saitoh E., Sanada K.;  
 CC RT "Fractionation and characterization of basic proline-rich peptides of  
 CC human parotid saliva and the amino acid sequence of proline-rich  
 CC peptide P-E.";  
 CC RL J. Biochem. 91:2067-2075(1982).  
 CC CC -!- MISCELLANEOUS: Peptides IB-9 and P-E are the same peptide.  
 CC DR PIR; B40750; PIHUB6.  
 CC DR GO; GO:0005576; C:extracellular; NAS.  
 CC KW Repeat; Parotid gland.  
 CC FT REPEAT 1 21 APPROXIMATE.  
 CC FT REPEAT 22 42 APPROXIMATE.  
 CC FT REPEAT 43 61 APPROXIMATE.  
 CC SQ SEQUENCE 61 AA; 6024 MW; F883F78617E05327 CRC64;  
 CC  
 CC Query Match 75.9%; Score 44; DB 1; Length 61;  
 CC Best Local Similarity 77.8%; Pred. No. 1.3;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 GRPQGPQQ 10  
 DB 4 GRPQGPQQ 12  
 CC  
 CC RESULT 9  
 CC PRPP\_HUMAN STANDARD; PRT; 96 AA.  
 CC ID PRPP\_HUMAN  
 CC AC P04281;  
 CC DT 20-MAR-1987 (Rel. 04, Created)  
 CC DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Basic proline-rich peptide IB-1.  
 CC OS Homo sapiens (Human)  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC OX NCBI\_TaxID=9606;  
 CC RN [1]  
 CC RP SEQUENCE  
 CC RX MEDLINE=86243355; PubMed=3521730;  
 CC RA Kauffman D., Hofmann T., Bannick A., Keller P.;  
 CC RT "Basic proline-rich proteins from human parotid saliva: complete  
 CC covalent structures of proteins IB-1 and IB-6.";  
 CC RL Biochemistry 25:2387-2392(1986).  
 CC CC -!- DOMAIN: Contains 21-residue repeats, two of which have internal  
 CC 7-residue repeats.  
 CC CC -!- MISCELLANEOUS: There are nine basic proline-rich peptides isolated  
 CC from the saliva; this peptide is designated IB-1.  
 CC DR PIR; B40750; PIHUB6.  
 CC KW Phosphorylation; Repeat; Parotid gland; Pyrrolidone carboxylic acid.  
 CC FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 CC FT MOD\_RES 8 8 PHOSPHORYLATION.  
 CC FT REPEAT 15 35 APPROXIMATE.  
 CC FT REPEAT 36 56 APPROXIMATE.  
 CC FT REPEAT 57 77 APPROXIMATE.  
 CC SQ SEQUENCE 96 AA; 9530 MW; FF3B3F68DE104F8 CRC64;  
 CC  
 CC Query Match 75.9%; Score 44; DB 1; Length 96;  
 CC Best Local Similarity 77.8%; Pred. No. 2.1;  
 CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC QY 2 GRPQGPQQ 10  
 CC DB 39 GRPQGPQQ 47  
 CC  
 CC RESULT 10  
 CC PRPP\_HUMAN STANDARD; PRT; 174 AA.  
 CC ID PRPP\_HUMAN  
 CC AC P81489;  
 CC DT 15-DEC-1998 (Rel. 37, Created)  
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Salivary proline-rich protein II-1 (Fragment).  
 CC OS Homo sapiens (Human)  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC OX NCBI\_TaxID=9606;  
 CC RN [1]  
 CC RP SEQUENCE  
 CC RC TISSUE=Saliva;  
 CC RX MEDLINE=93383383; PubMed=8373986;  
 CC RA Kaufman D.L., Keller P.J., Bannick A., Blum M.;  
 CC RT "Alignment of amino acid and DNA sequences of human proline-rich  
 CC proteins.";  
 CC RL Crit. Rev. Oral Biol. Med. 4:287-292(1993).  
 CC DR GO; GO:0005576; C:extracellular; NAS.  
 CC DR InterPro; IPR000637; AT\_hook.  
 CC DR PRINTS; PR00329; ATHOOK.  
 CC KW Repeat; Parotid gland; Multigene family.  
 CC FT NON\_TER 174 174



```

RESULT 12
ID      CES4_HUMAN          STANDARD;          PRT;    124 AA.
AC      Q9BXQ7;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DE      Cat eye syndrome critical region protein 4 (Fragment).
CE      CE04.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RS      SEQUENCE FROM N.A.
RR      MEDLINE=21275466; PubMed=11381032;
RA      Fotsz T.K., Brinkman-Mills P., Banting G.S., Maier S.A., Riazi M.A.,
RA      Bredtland L.J., Hu S., Birren B., Minoshima S., Shimizu N., Pan H.,
RA      Nguyen T., Fang F., Fu Y., Ray L., Wu H., Shaull S., Phan S., Yao Z.,
RA      Chen F., Huan A., Hu P., Wang Q., Loh P., Qi S., Roe B.A.,
RA      McDermid H.E.;
RA      Analysis of the cat eye syndrome critical region in humans and the
RA      region of conserved synteny in mice: a search for candidate genes at
RA      or near the human chromosome 22 pericentromere."
RRL      Genome Res. 11:1053-1070(2001).
CC      -!- TISSUE SPECIFICITY: Adult heart and skeletal muscle. Widely
CC      expressed in fetal tissues.
CC      -!- MISCELLANEOUS: Candidate gene for the Cat Eye Syndrome (CES), a
CC      developmental disorder associated with the duplication of a 2 Mb
CC      region of 22q11.2. Duplication usually takes in the form of a
CC      supernumerary bisatellited isodicentric chromosome, resulting in
CC      four copies of the region (represents an inv dup(22)(q11)). CES is
CC      characterized clinically by the combination of coloboma of the
CC      iris and anal atresia with fistula, downslanting palpebral
CC      fissures, preauricular tags and/or pits, frequent occurrence of
CC      heart and renal malformations, and normal or near-normal mental
CC      development.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AF307448; AAK30048.1; -.
CC      Genew: HGNC:1842; CE04.
CC      NON TER
CC      1
CC      SEQUENCE 124 AA; 12838 MW; 938E00386308EC7A CRC64;
CC      Query Match 72.4%; Score 42; DB 1; Length 124;
CC      Best Local Similarity 87.5%; Pred. No. 5.4;
CC      Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC      QY 1 RGRPQGP 8
CC      |||||
CC      58 RGRPQAPP 65
CC      DB
RESULT 13
SVB_LQPE
ID      SVB_LQPE          STANDARD;          PRT;    125 AA.
AC      P47194;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Synapcobrevin.
DE      Lolligo pealeii (Longfin squid).
OC      Eukaryota; Metazoa; Mollusca; Cephalopoda; Colecoidea; Neocoleoidea;
OC      Decapodiformes; Lolliginidae; Lolligo.

```

```

OX NCBI_TaxID=6621;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94280860; PubMed=8011337;
RA Hunt J.M., Bomert K., Charlton M.P., Kistner A., Habermann E.,
RA Augustine G.J., Betz H.;
RT "A post-docking role for synaptobrevin in synaptic vesicle fusion.";
RL Neuron 12:1269-1279(1994).
CC -!- FUNCTION: Intrinsic membrane protein of small synaptic
CC vesicles.
CC -!- SUBCELLULAR LOCATION: Type IV membrane protein. Neuronal synaptic
CC vesicles.
CC -!- SIMILARITY: Belongs to the synaptobrevin family.
CC -!- SIMILARITY: Contains 1 v-SNARE coiled-coil homology domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X74748; CAA52766.1; -.
CC PIR; S40153; S40153.
CC PDB; 1L4A; 31-JUL-02.
CC InterPro; IPR001388; Synaptobrevin.
CC Pfam; PF00957; synaptobrevin; 1.
CC PRINTS; PR00219; SYNAPTOBREVN.
CC ProDom; PD001229; Synaptobrevin; 1.
CC PROSITE; PS00417; SYNAPTOBREVN; 1.
CC PROSITE; PS08982; V-SNARE; 1.
CC Synapse; Synaptosome; Transmembrane; Coiled coil; 3D-structure.
KW DOMAIN 1 103 CYTOPLASMIC (POTENTIAL).
FT ANCHOR FOR TYPE IV MEMBRANE PROTEIN
FT TRANSMEM 104 123 (POTENTIAL).
FT DOMAIN 124 125 VESICULAR (POTENTIAL).
FT DOMAIN 40 100 V-SNARE COILED-COIL HOMOLGY.
SQ SEQUENCE 125 AA; 13316 MW; 6FE9B5604ECABD86 CRC64;
Query Match 70.7%; Score 41; DB 1; Length 125;
Best Local Similarity 87.5%; Pred. No. 7.7;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 GRPQGPQP 9
Db 26 GPPQGPQP 33
RESULT 14
SCAP_CRIGR STANDARD; PRT; 1276 AA.
AC P97260;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sterol regulatory element binding protein cleavage-activating protein
DE (SREBP cleavage-activating protein) (SCAP).
GN SCAP.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF ASP-443.
RX TISSUE=Ovary;
RA MEDLINE=97053781; PubMed=8998195;
RA Hua X., Nonturff A., Goldstein J.L., Brown M.S.;
RT "Sterol resistance in CHO cells traced to point mutation in SREBP
RT cleavage-activating protein.";
RL Cell 87:415-426(1996).
RN [2]

```

```

RP POST-TRANSLATIONAL MODIFICATIONS, AND MUTAGENESIS OF TYR-298.
RC TISSUE=Ovary;
RX MEDLINE=99432216; PubMed=10500160;
RA Nonturff A., DeRose-Boyd R.A., Cheek S., Goldstein J.L., Brown M.S.;
RT "Sterols regulate cycling of SREBP cleavage-activating protein (SCAP)
RT between endoplasmic reticulum and Golgi.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11235-11240(1999).
RN [3]
RP TOPOLOGY, AND CARBOHYDRATE-LINKAGE SITES ASN-263; ASN-590 AND ASN-641.
RX MEDLINE=98307971; PubMed=9642295;
RA Nonturff A., Brown M.S., Goldstein J.L.;
RT "Topology of SREBP cleavage-activating protein, a polytopic membrane
RT protein with a sterol-sensing domain.";
RL J. Biol. Chem. 273:17243-17250(1998).
RN [4]
RP SUBCELLULAR LOCATION.
RX MEDLINE=20449077; PubMed=10896675;
RA Yang T., Goldstein J.L., Brown M.S.;
RT "Overexpression of membrane domain of SCAP prevents sterols from
RT inhibiting SCAP. SREBP exit from endoplasmic reticulum.";
RL J. Biol. Chem. 275:29881-29886(2000).
CC -!- FUNCTION: Sterol sensor. Necessary for the proteolytic activation
CC of SREBPs by site-1 protease in the Golgi.
CC -!- SUBUNIT: In a tight complex with SREBs.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Moves from the
CC endoplasmic reticulum to the Golgi in the absence of sterols.
CC -!- SIMILARITY: Contains 1 sterol sensor (SSD) domain.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67060; AAB19103.1; -.
CC PIR; T18526; T18526.
CC GO; GO:0005783; C:endoplasmic reticulum; ISS.
CC GO; GO:0005794; C:Golgi apparatus; ISS.
CC GO; GO:0016021; C:integral to membrane; ISS.
CC GO; GO:0003754; F:chaperone activity; NAS.
CC GO; GO:0045541; F:negative regulation of cholesterol biosynth. . ; ISS.
CC GO; GO:0045716; P:positive regulation of low-density lipoprot. . ; ISS.
CC GO; GO:0006994; P:sterol depletion response, SREBP target gen. . ; ISS.
CC InterPro; IPR000731; SSD 5TM.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 4.
CC SMART; SM00320; WD40; 6.
CC PROSITE; PS0156; SSD; 1.
CC PROSITE; PS00678; WD REPEATS_1; 1.
CC PROSITE; PS00678; WD REPEATS_2; 1.
CC PROSITE; PS00682; WD REPEATS_2; 1.
CC PROSITE; PS0294; WD REPEATS REGION; 1.
KW Lipid metabolism; Cholesterol metabolism; Transmembrane;
KW Endoplasmic reticulum; Golgi stack; Repeat; WD repeat; Glycoprotein.
FT DOMAIN 1 18 CYTOPLASMIC.
FT TRANSMEM 19 39 1 (POTENTIAL).
FT DOMAIN 40 279 LUMENAL.
FT TRANSMEM 280 300 2 (POTENTIAL).
FT DOMAIN 301 312 CYTOPLASMIC.
FT TRANSMEM 313 333 3 (POTENTIAL).
FT DOMAIN 334 344 LUMENAL.
FT TRANSMEM 345 365 4 (POTENTIAL).
FT DOMAIN 366 401 CYTOPLASMIC.
FT TRANSMEM 402 422 5 (POTENTIAL).
FT DOMAIN 423 423 LUMENAL.
FT TRANSMEM 424 444 6 (POTENTIAL).
FT DOMAIN 445 518 CYTOPLASMIC.
FT TRANSMEM 519 539 7 (POTENTIAL).
FT DOMAIN 540 708 LUMENAL.
FT TRANSMEM 709 729 8 (POTENTIAL).
FT DOMAIN 730 1276 CYTOPLASMIC.

```

Search completed: April 6, 2004, 16:07:59  
Job time : 6.4486 secs

```
FT DOMAIN 284 442 SSD.  
FT REPEAT 771 811 WD 1.  
FT REPEAT 949 999 WD 2.  
FT REPEAT 1002 1039 WD 3.  
FT REPEAT 1074 1111 WD 4.  
FT REPEAT 1114 1152 WD 5.  
FT REPEAT 1155 1192 WD 6.  
FT REPEAT 1194 1232 WD 7.  
FT DOMAIN 722 729 POLY-LEU.  
FT DOMAIN 747 750 POLY-ARG.  
FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 641 641 Y->C: ABOLISHES STEROL-RESPONSE.  
FT MUTAGEN 298 298 D->N: ABOLISHES STEROL-RESPONSE.  
FT MUTAGEN 443 443  
SQ SEQUENCE 1276 AA; 139512 MW; A8693F7157FF5FFC CRC64;  
Query Match 70.7%; Score 41; DB 1; Length 1276;  
Best Local Similarity 77.8%; Pred. No. 76;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Cy 1 RGRPGQPPQ 9  
Db 854 RHRPGGPPQ 862
```

```
RESULT 15  
PRP2_RAT STANDARD; PRT; 172 AA.  
AC P10164;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Acidic proline-rich protein PRP25 precursor (Fragment).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Parotid gland;  
RX MEDLINE=86033799; PubMed=3840480;  
RA Clements S.; Mehansho H.; Carlson D.M.;  
RT "Novel multigene families encoding highly repetitive peptide  
RT sequences. Sequence analyses of rat and mouse proline-rich protein  
RT CDNAS.";  
RL J. Biol. Chem. 260:13471-13477(1985).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; M11901; AAA41948.1; --  
DR PIR; B29149; B29149.  
KW Repeat; Parotid gland; Multigene family; Signal.  
FT SIGNAL 1 16 POTENTIAL.  
FT CHAIN 17 >172 ACIDIC PROLINE-RICH PROTEIN PRP25.  
FT NON TER 172 172  
SQ SEQUENCE 172 AA; 17416 MW; F63BFD0545D6EA CRC64;  
Query Match 69.0%; Score 40; DB 1; Length 172;  
Best Local Similarity 85.7%; Pred. No. 15;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 GRPGQPP 8  
Db 85 GKPGQPP 91
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 36.1682 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-6  
Perfect score: 58  
Sequence: 1 GRPQGPQQ 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25.\*

1: sp\_arched.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	86.2	93	Q15215	Q15215 homo sapien
2	50	86.2	173	Q15214	Q15214 homc sapien
3	50	86.2	188	Q29427	Q29427 macaca fasc
4	50	86.2	238	Q00600	Q00600 homo sapien
5	50	86.2	297	Q16038	Q16038 homo sapien
6	50	86.2	338	Q86VAL	Q86VAL homo sapien
7	50	86.2	358	Q08805	Q08805 homo sapien
8	50	86.2	382	Q00599	Q00599 homo sapien
9	48	82.8	256	Q7UYG7	Q7UYG7 rhodopirell
10	48	82.8	295	Q10761	Q10761 rattus norv
11	47	81.0	796	Q8NKJ3	Q8NKJ3 saccharomyc
12	45	77.6	1559	Q86BS0	Q86BS0 drosophila
13	45	77.6	1569	Q9VF66	Q9VF66 drosophila
14	45	77.6	1622	Q86BP0	Q86BP0 drosophila
15	44	75.9	46	Q15218	Q15218 homo sapien
16	44	75.9	170	Q07610	Q07610 rattus norv

17	44	75.9	309	4	Q04118
18	44	75.9	395	11	Q9JIB1
19	43	74.1	188	11	Q62106
20	43	74.1	317	11	Q62103
21	43	74.1	333	4	Q96SD4
22	43	74.1	367	4	Q8N6Q2
23	43	74.1	367	4	Q8N6Q2
24	43	74.1	367	4	Q8N6Q2
25	43	74.1	645	5	Q9U302
26	43	74.1	645	5	Q17350
27	43	74.1	803	16	Q812W2
28	43	74.1	1089	5	Q26155
29	42	72.4	85	3	Q9VR11
30	42	72.4	109	3	Q9P8A8
31	42	72.4	204	16	Q7U416
32	42	72.4	339	13	Q7ZUX2
33	42	72.4	504	10	Q9LH00
34	42	72.4	504	11	Q91X93
35	41	70.7	18	4	Q9UCT9
36	41	70.7	220	13	Q7T339
37	41	70.7	395	11	Q91Z76
38	41	70.7	395	11	Q8BZ39
39	41	70.7	579	2	Q8GRF2
40	41	70.7	621	2	Q7WSY2
41	41	70.7	771	11	Q8C018
42	41	70.7	789	11	Q8C038
43	41	70.7	807	11	Q91Y59
44	41	70.7	809	11	Q7TS83
45	41	70.7	891	5	Q9V338

## ALIGNMENTS

## RESULT 1

Q15215 PRELIMINARY; PRT; 93 AA.  
AC Q15215  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Salivary proline-rich protein 1 (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84298176; PubMed=6089212;  
RA Azen E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S., Maeda N., Vanin E.F., Carlson D.M., Smithies O.;  
RT "Clones from the human gene complex coding for salivary proline-rich proteins";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565(1984).  
DR EMBL; K02576; AAA36503.1; --  
FT NON\_TER 1  
SQ SEQUENCE 93 AA; 9170 MW; 04AD5F0797E31867 CRC64;

Query Match 86.2%; Score 50; DB 4; Length 93;  
Best Local Similarity 88.9%; Pred. No. 0.42;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10  
DB 41 GRPQGPQQ 49

## RESULT 2

Q15214 PRELIMINARY; PRT; 173 AA.  
ID Q15214  
AC Q15214  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Salivary proline-rich protein 1 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8429176; PubMed=6089212;  
 RA Azen E.A., Lyons K.M., McConigal T., Bartlett N.L., Clements L.S.,  
 RA Macdonald N., Vanin E.F., Carlson D.M., Smithies O.,  
 RT "Clones from the human gene complex coding for salivary proline-rich  
 RT proteins";  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565(1984).  
 DR EMBL; K02575; AAA36502.1; -;  
 DR GO; GO:0005576; C:extracellular; NAS.  
 FT NON\_TER 1 1  
 FT NON\_TER 173 173  
 SQ SEQUENCE 173 AA; 17206 MW; BB4E30335B7A7C0 CRC64;  
  
 Query Match 86.2%; Score 50; DB 4; Length 173;  
 Best Local Similarity 88.9%; Pred. No. 0.79;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 2 GRPQGPQQ 10  
 DB 160 GKPGPPQQ 168  
  
 RESULT 3  
 Q29427 PRELIMINARY; PRT; 188 AA.  
 AC Q29427  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Proline-rich protein.  
 GN MNP4 OR MNPFP9.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92009216; PubMed=1916292;  
 RA Lin H.H., Kousvelari E.E., Ann D.K.;  
 RT "Sequence and expression of the Mnp4 gene encoding basic proline-rich  
 RT protein in macaque salivary glands";  
 RL Gene 104:219-226(1991).  
 DR EMBL; M81322; AAA36905.1; -;  
 DR EMBL; M81321; AAA36904.1; -;  
 DR FIR; JH0481; JH0481.  
 SQ SEQUENCE 188 AA; 19135 MW; C9E6D0B7F4DEES04 CRC64;  
  
 Query Match 86.2%; Score 50; DB 6; Length 188;  
 Best Local Similarity 88.9%; Pred. No. 0.86;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 2 GRPQGPQQ 10  
 DB 95 GKPGPPQQ 103  
  
 RESULT 4  
 Q00600 PRELIMINARY; PRT; 238 AA.  
 AC Q00600  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Parotid 'o' protein (Fragment).  
 GN PRB4.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96108975; PubMed=8554050;  
 RA Azen E.A., Amberger E., Fisher S., Prakobphol A., Niece R.L.;  
 RT "PRB1, PRB2, and PRB4 coded polymorphisms among human salivary  
 RT concanavalin-A binding, II-1, and Po proline-rich proteins";  
 RL Am. J. Hum. Genet. 58:143-153(1996).  
 DR EMBL; S80916; AAB50687.2; -;  
 FT NON\_TER 1 1  
 SQ SEQUENCE 238 AA; 24005 MW; 4F44E947FF3A6C1 CRC64;  
  
 Query Match 86.2%; Score 50; DB 4; Length 238;  
 Best Local Similarity 88.9%; Pred. No. 1.1;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 2 GRPQGPQQ 10  
 DB 172 GKPGPPQQ 180  
  
 RESULT 5  
 Q16038 PRELIMINARY; PRT; 297 AA.  
 AC Q16038  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE PRB1M protein (Fragment).  
 GN PRB1M.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93304421; PubMed=8317492;  
 RA Azen E.A., Latreille P., Niece R.L.;  
 RT "PRB1 gene variants coding for length and null polymorphisms among  
 RT human salivary Pg, PmF, PmS, and Pe proline-rich proteins (PRPs)";  
 RL Am. J. Hum. Genet. 53:264-278(1993).  
 DR EMBL; S62928; AAB27288.2; -;  
 DR FIR; B40750; FIKUB6.  
 DR FIR; D40750; D40750.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 297 AA; 29046 MW; 1C7BFE4CAGB5B5F0 CRC64;  
  
 Query Match 86.2%; Score 50; DB 4; Length 297;  
 Best Local Similarity 88.9%; Pred. No. 1.4;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 2 GRPQGPQQ 10  
 DB 143 GKPGPPQQ 151  
  
 RESULT 6  
 Q86YAL PRELIMINARY; PRT; 338 AA.  
 AC Q86YAL  
 DT 01-JUN-2003 (TREMBLrel. 24, Created)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC044827; AAH44827.1; -.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 338 AA; 33344 MW; 7F54B4E5AB002261 CRC64;  
Query Match 86.2%; Score 50; DB 4; Length 338;  
Best Local Similarity 88.9%; Pred. No. 1.6;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GRPQGPQQ 10  
Db 184 GRPQGPQQ 192  
RESULT 7  
Q08805 PRELIMINARY; PRT; 358 AA.  
AC Q08805;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE PRB1L protein (Fragment).  
GN PRB1L.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93304421; PubMed=8317492;  
RA Azen E.A., Latreille P., Niece R.L.;  
RT "PRB1 gene variants coding for length and null polymorphisms among  
human salivary Ps, Pn2P, Pms, and Pn proline-rich proteins (PRPs).";  
RL Am. J. Hum. Genet. 53:264-278(1993).  
DR EMBL; S62941; AAB27289.1; -.  
DR FIR; D40750; D40750.  
FT NON\_TER 1  
SQ SEQUENCE 358 AA; 35050 MW; DB7F87B8D5EA759E CRC64;  
Query Match 86.2%; Score 50; DB 4; Length 358;  
Best Local Similarity 88.9%; Pred. No. 1.7;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GRPQGPQQ 10  
Db 204 GRPQGPQQ 212  
RESULT 8  
Q00599 PRELIMINARY; PRT; 382 AA.  
AC Q00599;  
DT 01-JUL-1997 (TRENBLrel. 04, Created)  
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Con1 (Fragment).  
GN PRB2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96108975; PubMed=8554050;  
RA Azen E.A., Amberger E., Fisher S., Prakhobphol A., Niece R.L.;  
RT "PRB1, PRB2, and PRB4 coded polymorphisms among human salivary  
concanavalin-A binding, II-1, and Pn proline-rich proteins.";  
RL Am. J. Hum. Genet. 58:143-153(1996).  
DR EMBL; S80905; AAB50686.1; -.  
FT NON\_TER 1

SQ SEQUENCE 382 AA; 37277 MW; D500942AB979209D CRC64;  
Query Match 86.2%; Score 50; DB 4; Length 382;  
Best Local Similarity 88.9%; Pred. No. 1.8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GRPQGPQQ 10  
Db 330 GRPQGPQQ 338  
RESULT 9  
Q7UYG7 PRELIMINARY; PRT; 256 AA.  
AC Q7UYG7;  
DT 01-OCT-2003 (TRENBLrel. 25, Created)  
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN RB624.  
OS Rhodopirellula baltica.  
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
OC Planctomycetaceae; Pirellula.  
OX NCBI\_TaxID=117;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=1;  
RX MEDLINE=22735913; PubMed=12835416;  
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
RA Schlesner H., Amann R., Reinhardt R.;  
RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
strain 1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
RW EMBL; BX294134; CAD71675.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 256 AA; 27003 MW; 2FAA06B5206E8B2F CRC64;  
Query Match 82.8%; Score 48; DB 16; Length 256;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRQGP 8  
Db 160 RGRQGP 167  
RESULT 10  
Q07611 PRELIMINARY; PRT; 295 AA.  
AC Q07611;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Proline-rich proteoglycan.  
GN PRPG2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Sprague-Dawley; Tissue=Parotid gland;  
RX MEDLINE=93388826; PubMed=8376404;  
RA Castle A.M., Castle J.D.;  
RT "Novel secretory proline-rich proteoglycans from rat parotid. Cloning  
and characterization by expression in Att-20 cells.";  
RL J. Biol. Chem. 268:20490-20496(1993).  
DR EMBL; L17318; AAA03074.1; -.  
DR FIR; B48013; B48013.  
SQ SEQUENCE 295 AA; 30026 MW; EEA44BD8B3BE1B9F CRC64;  
Query Match 82.8%; Score 48; DB 11; Length 295;



RC STRAIN=S288C;  
 RA Gromadka R.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [16]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA MIPS;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X59720; CAC42990.1; -;  
 DR SGD; S0000673; PAT1.  
 DR GO; GO:000509; F:calcium ion binding; IEA.  
 DR GO; GO:0016853; F:isomerase activity; IEA.  
 DR InterPro; IPR002048; EF-hand; 1.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW isomerase.  
 SQ SEQUENCE 796 AA; 88495 MW; 8DBE0DEF36D289DE CRC64;

Query Match 81.0%; Score 47; DB 3; Length 796;  
 Best Local Similarity 80.0%; Pred. No. 12;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGGPQQ 10  
 DB 168 QGLPGGPQQ 177

RESULT 12  
 Q86BS0 PRELIMINARY; PRT; 1559 AA.  
 AC Q86BS0;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CG31302-PB.  
 GN CG31302  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams W.D., Ceinliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith I.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

SVirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C.M., Berman J.P., Carlson J.W., Ceinliker S.E.,  
 RA Ciamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,  
 RA Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,  
 Russo S., Searle S.M.J., Smith E., Shu S., Smutniak P.,  
 RA Whitfield E.J., Ashburner M., Gelbart W.M., Mungall C.J.,  
 Lewis S.E.;  
 RA "Annotation of Drosophila melanogaster genome.";  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003708; AAP55194.3; -;  
 DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.  
 DR GO; GO:0003774; F:motor activity; IEA.  
 DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
 DR InterPro; IPR000563; Flag\_FliH.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00041; fn3; 3.  
 DR Pfam; PF00018; SH3; 2.  
 DR PRINTS; PR01003; FLGFLIH.  
 DR ProDom; PD000066; SH3; 2.  
 DR SMART; SMO0060; FN3; 3.  
 DR SMART; SMO0326; SH3; 3.  
 DR PROSITE; PS00002; SH3; 3.  
 SQ SEQUENCE 1559 AA; 163964 MW; 255325F16B70697D CRC64;  
 Query Match 77.6%; Score 45; DB 5; Length 1559;  
 Best Local Similarity 88.9%; Pred. No. 52;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPGGPQQ 10

DB 866 GRPGGPQQ 874

RESULT 13

Q9VF66

ID Q9VF66 PRELIMINARY; PRT; 1569 AA.

AC Q9VF66;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE CG31302-PA.

GN CG31302

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams W.D., Ceinliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,



George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L., Abril J.P., Agayani A., An H.J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W., Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Li X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobery C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Muzny D.M., Nelson D.L., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhang W., Zhang G., Zhao Q., Zhou X., Zhu S., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of *Drosophila melanogaster*.";  
 Science 287:2185-2195(2000).  
 [2]  
 SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K., Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D., Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celisner S.E., Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J., Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J., Russo S., Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

"Annotation of *Drosophila melanogaster* genome.";  
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 SEQUENCE FROM N.A.  
 RA FlyBase;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RA FlyBase;  
 RA Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RA EMBL; A0003708; AAF55193.3;  
 DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.  
 DR GO; GO:0003774; F:motor activity; IEA.  
 DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
 DR InterPro; IPR000563; Flag Flah.  
 DR InterPro; IPR003961; FN-III.  
 DR InterPro; IPR008957; FN-III-like.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00041; FN3; 3.  
 DR Pfam; PF00018; SH3; 2.  
 DR PRINTS; PR01003; FLGFLTH.  
 DR PRODOM; PD000066; SH3; 2.  
 DR SMART; SM00060; FN3; 3.  
 DR SMART; SM00326; SH3; 3.  
 DR PROSITE; PS00002; SH3; 3.  
 SQ SEQUENCE 1569 AA; 165151 MW; 81A008F054B07FFF CRC64;

Query Match 77.6%; Score 45; DB 5; Length 1569;  
 Best Local Similarity 88.9%; Pred. No. 52;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GRPQGPQQ 10  
 |||||  
 DB 876 GRPQGPQQ 884  
 |||||

RESULT 14  
 Q86BP0 PRELIMINARY; PRT; 1622 AA.  
 ID Q86BP0;  
 AC Q86BP0;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CG31302-PC.  
 GN CG31302.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RW [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celisner S.E., Li P.W., Hoskins R.A., Galle R.F., Amanatides P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N., George R.A., Lewis S.E., Yandell M.D., Zhang Q., Chen L.X., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers J.H., Blazek R.G., Champagne M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L., Abril J.P., Agayani A., An H.J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W., Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Li X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobery C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Muzny D.M., Nelson D.L., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhang W., Zhang G., Zhao Q., Zhou X., Zhu S., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of *Drosophila melanogaster*.";  
 Science 287:2185-2195(2000).  
 [2]  
 SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K., Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D., Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celisner S.E., Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J., Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J., Russo S., Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

"Annotation of *Drosophila melanogaster* genome.";  
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 SEQUENCE FROM N.A.  
 RA FlyBase;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RA FlyBase;  
 RA Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RA EMBL; A0003708; AAF55193.3;  
 DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.  
 DR GO; GO:0003774; F:motor activity; IEA.  
 DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.  
 DR InterPro; IPR000563; Flag Flah.  
 DR InterPro; IPR003961; FN-III.  
 DR InterPro; IPR008957; FN-III-like.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00041; FN3; 3.  
 DR Pfam; PF00018; SH3; 2.  
 DR PRINTS; PR01003; FLGFLTH.  
 DR PRODOM; PD000066; SH3; 2.  
 DR SMART; SM00060; FN3; 3.  
 DR SMART; SM00326; SH3; 3.  
 DR PROSITE; PS00002; SH3; 3.  
 SQ SEQUENCE 1569 AA; 165151 MW; 81A008F054B07FFF CRC64;

RA Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Flybase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Flybase;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A0003708; AAN13660.2; ..  
 DR GO; GO:0003288; Cytoskeleton (sensu Bacteria); IEA.  
 DR GO; GO:0003774; F-motor activity; IEA.  
 DR GO; GO:0001539; P-ciliary/flagellar motility; IEA.  
 DR InterPro; IPR000563; Flag\_FliH.  
 DR InterPro; IPR003961; FN III.  
 DR InterPro; IPR008957; FN III-like.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00041; fn3; 3.  
 DR Pfam; PF00018; SH3; 2.  
 DR PRINTS; PR01003; FLGELIH.  
 DR PRODOM; PD000066; SH3; 2.  
 DR SMART; SM00060; FN3; 3.  
 DR SMART; SM00326; SH3; 3.  
 DR PROSITE; PS00002; SH3; 3.  
 SQ SEQUENCE 1622 AA; 170827 MW; 8C98EB8D3B307B3A CRC64;

Query Match 77.6%; Score 45; DB 5; Length 1622;  
 Best Local Similarity 88.9%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10  
 |||||  
 Db 929 GRPQGPQQ 937

## RESULT 15

Q15218 Q15218 PRELIMINARY; PRT; 46 AA.  
 AC Q15218;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)  
 DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)  
 DE Salivary proline-rich protein 2 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84298176; PubMed=6089212;  
 RA Aren E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S.,  
 RA Maeda N., Vanin E.F., Carlson D.M., Smithies O.;  
 RT "Clones from the human gene complex coding for salivary proline-rich  
 RT proteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565(1984).  
 DR EMBL; K02578; AAA36505.1; ..  
 DR NON\_TER 1 1  
 FT NON\_TER 45 46  
 SQ SEQUENCE 46 AA; 4592 MW; FCE1D38D8DEDC173 CRC64;

Query Match 75.9%; Score 44; DB 4; Length 46;  
 Best Local Similarity 87.5%; Pred. No. 2;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RPOGPPQQ 10  
 :|||  
 Db 8 KPOGPPQQ 15

Search completed: April 6, 2004, 16:14:40  
 Job time : 37.2451 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 55.9813 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-6  
Perfect score: 58  
Sequence: 1 RGRPPQPPQQ 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	100.0	10	4	AAB48776 Human sal
2	58	100.0	17	4	AAB48783
3	58	100.0	132	4	ABP38848 Peptide #
4	58	100.0	132	4	AAM32323 Peptide #
5	58	100.0	132	4	AAM72058 Human bon
6	58	100.0	132	4	AAM59494 Human bra
7	58	100.0	132	4	ABG53742 Human liv
8	58	100.0	132	5	ABG41873 Human pep
9	58	100.0	149	6	ABP57423 Human NOV
10	58	100.0	154	6	ABP57423 Human sec
11	58	100.0	166	6	ADAB3798 Human PRH
12	58	100.0	166	7	ADC98216 Human sal
13	53	91.4	9	4	AAB48775
14	53	91.4	44	2	AAW03557 Human pro
15	53	91.4	45	1	APB93320 P-C Gene
16	51	87.9	106	6	ABP76137 Human GEN
17	51	87.9	106	6	ABP75971 Human GEN
18	50	86.2	117	6	AAO30249 Human Sap
19	50	86.2	192	6	AAO30246 Human Sap
20	50	86.2	208	6	AAO30188 Human cp3
21	50	86.2	247	6	ABU11891 Human ABC
22	50	86.2	331	7	ADD45740 Human Pro
23	48	82.8	8	4	AAB48774
24	48	82.8	111	6	ABP75970 Human GEN
25	48	82.8	111	6	ABP76136 Human GEN

26	48	82.8	296	7	ADD45442	Rat Prote
27	48	82.8	296	7	ADP57105	Rat Prote
28	47	81.0	796	6	ABP53413	Protein s
29	45	77.6	266	4	ABG05763	Novel hum
30	45	77.6	1251	4	ABB61254	ABG05763
31	44	75.9	82	2	AAW50193	Amino aci
32	44	75.9	124	7	AAW50192	Amino aci
33	44	75.9	162	7	AAO30407	Human sec
34	44	75.9	267	7	AAO30408	Human sec
35	44	75.9	395	4	AAO30364	Rat G-pro
36	44	75.9	395	5	AAE14263	Rat NMUR2
37	44	75.9	591	4	ABG15767	Novel hum
38	44	75.9	591	4	ABG18110	Novel hum
39	43	74.1	112	4	AAO02189	Human pol
40	43	74.1	667	7	ADD67376	FLAG epit
41	42	72.4	49	3	ABP34155	Gene 13 h
42	42	72.4	85	4	ABP58372	Drosophil
43	42	72.4	1146	6	ABP34447	Human lip
44	41	70.7	7	4	AAB48773	Human sal
45	41	70.7	104	4	AAU50351	Propionib

## ALIGNMENTS

RESULT 1  
AAB48776  
ID AAB48776 standard; peptide; 10 AA.

XX AAB48776;  
AC AAB48776;  
DT 09-MAY-2001 (first entry)  
XX Human saliva PRP-1 fragment (residues 106-115), SEQ ID NO:6.  
KW Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX Homo sapiens.  
XX WO200069890-A1.  
XX 23-NOV-2000.  
PD 23-NOV-2000.  
PF 11-MAY-2000; 2000WO-SE000930.  
XX 17-MAY-1999; 99SE-00001773.  
PR 17-MAY-1999; 99SE-00001773.  
XX (STRO)/ STROEMBERG N.  
XX (JOHA)/ JOHANSSON I.  
XX Stroemberg N, Johansson I;  
XX WPI; 2001-031923/04.  
PT New oligopeptides comprising 2 arginine residues from degradation of  
PT proline-rich proteins, useful for preventing dental caries.  
PS Claim 4; Page 24; 36pp; English.

The invention relates to human PRP-1-derived oligopeptides (AAB48771-AAB48783) which contain at least two arginine residues and which protect against dental caries. PRPs (proline-rich proteins) are salivary proteins encoded by six clustered genes on chromosome 12p13.2 and are potential determinants of a person's susceptibility to dental caries. PRPs are degraded by Actinomyces and Streptococcus species to small peptide fragments. These are metabolised by oral bacteria for nutritional purposes, with certain bacterial species generating ammonia via the catabolism of arginine. The peptides of the invention, being arginine-rich, can also be converted to ammonia by these bacteria. The ammonia thus formed raises the pH at the dental surface, thereby protecting the teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

CC derived oligopeptides of the invention

SQ Sequence 10 AA;

Query Match 100.0%; Score 58; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.026;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPGQPQQ 10  
| | | | |  
Db 1 RGRPGQPQQ 10

RESULT 2

AA48783  
ID AAB48783 standard; peptide; 17 AA.

XX AC

XX AAB48783;

XX 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;

XX XW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
oral bacterium; caries prevention.

XX XW

XX OS Homo sapiens.

XX XW WO200069890-A1.

XX PN 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SE000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX PA (STRO/) STROEMBERG N.  
(JOHA/) JOHANSSON I.

XX FI Stroemberg N, Johansson I;  
XX DR WPI; 2001-031923/04.

XX PT New oligopeptides comprising 2 arginine residues from degradation of  
proline-rich proteins, useful for preventing dental caries.

XX FS Claim 2; Page 24; 36pp; English.

XX CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
AAB48783) which contain at least two arginine residues and which protect  
against dental caries. PRPs (proline-rich proteins) are salivary proteins  
encoded by six clustered genes on chromosome 12p13.2 and are potential  
determinants of a person's susceptibility to dental caries. PRPs are  
degraded by Actinomyces and Streptococcus species to small peptide  
fragments. These are metabolised by oral bacteria for nutritional  
purposes, with certain bacterial species generating ammonia via the  
catabolism of arginine. The peptides of the invention, being arginine-  
rich, can also be converted to ammonia by these bacteria. The ammonia  
thus formed raises the pH at the dental surface, thereby protecting the  
teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
derived oligopeptides of the invention

XX SQ Sequence 17 AA;

Query Match 100.0%; Score 58; DB 4; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.042;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPGQPQQ 10  
| | | | |  
Db 8 RGRPGQPQQ 17

RESULT 3

AB38848

ID ABB38848 standard; peptide; 132 AA.

XX AC ABB38848;

XX 04-FEB-2002 (first entry)

XX DE Peptide #6354 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000669.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 03-JUN-2000; 2000US-00608409.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234587P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing  
gene expression in human fetal liver.

XX PS Claim 27; SEQ ID NO 31483; 639pp + Sequence Listing; English.

XX CC The invention relates to a single exon nucleic acid probe for measuring  
human gene expression in a sample derived from human foetal liver. The  
single exon nucleic acid probes may be used for predicting, measuring and  
displaying gene expression in samples derived from human fetal liver. The  
present sequence is a peptide encoded by a single exon nucleic acid probe  
of the invention. Note: The sequence data for this patent did not form  
part of the printed specification, but was obtained in electronic format  
directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 58; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 0.27;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPGQPQQ 10  
| | | | |  
Db 88 RGRPGQPQQ 97

RESULT 4

AA32323

ID AAM32323 standard; protein; 132 AA.

XX AC AAM32323;

XX 17-OCT-2001 (first entry)

XX DE Peptide #6360 encoded by probe for measuring placental gene expression.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;  
Genetic disorder.



CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 58; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 0.27;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGPQQ 10  
 DB 88 RGRPQGPQQ 97

## RESULT 7

ABG53742  
 ID ABG53742 standard; peptide; 132 AA.

XX AC ABG53742;

XX DT 25-FEB-2003 (first entry)

XX DE Human liver peptide, SEQ ID No 32390.

XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.

XX OS Homo sapiens.

XX PN WO200157273-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000664.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488898/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human adult liver.

XX PS Claim 27; SEQ ID NO 32390; 658pp; English.

XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
 CC liver single exon encoded peptides of the invention. Note: The sequence  
 CC information for this patent does not appear in the printed specification  
 CC but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 58; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 0.27;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRPQGPQQ 10  
 DB 88 RGRPQGPQQ 97

## RESULT 8

ABG41873  
 ID ABG41873 standard; peptide; 132 AA.

XX AC ABG41873;

XX DT 19-AUG-2002 (first entry)

XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 31538.

XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.

XX OS Homo sapiens.

XX PN WO200186003-A2.

XX PD 15-NOV-2001.

XX PF 30-JAN-2001; 2001WO-US000665.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2002-114183/15.

XX PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples.

XX PS Claim 27; SEQ ID NO 31538; 634pp; English.

XX CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of probes  
 CC; the novel set of probes which hybridise at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,

comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIFO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 132 AA;

Query Match 100.0%; Score 58; DB 5; Length 132;  
Best Local Similarity 100.0%; Pred. No. 0.27; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0;

QY 1 RGRPGQPQQ 10  
DB 88 RGRPGQPQQ 97

RESULT 9  
ABR57423  
ID ABR57423 standard; protein; 149 AA.

AC ABR57423;

DT 15-SEP-2003 (first entry)

DE Human NOV7 protein SEQ ID NO:24.

KW Human; NOVX; cytostatic; cardiac; antiinflammatory; immunosuppressive;  
KW antiallergic; haemostatic; anti-HIV; antidiabetic; antiarteriosclerotic;  
KW anorectic; antiasthmatic; nephrotropic; antiarthritic; hepatotropic;  
KW neuroprotective; nontropic; antibacterial; virucide; antiparasitic;  
KW relaxant; anticonvulsant; hypotensive; vasotropic; antiparkinsonian;  
KW vulnerary; angiogenic; antiangiogenic; gene therapy; vaccine; cancer;  
KW cardiomyopathy; atherosclerosis; hypertension; diabetes; inflammation;  
KW autoimmune disorder; allergy; blood disorder; AIDS; obesity; asthma;  
KW acquired immunodeficiency syndrome; nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; Parkinson's disease; goitre; infection; stroke;  
KW muscular dystrophy; epilepsy; wasting disorder; chromosome 12.

OS Homo sapiens.

PN WO200294870-A2.

PD 28-NOV-2002.

PF 02-NOV-2001; 2001WO-US051580.

XX 02-NOV-2000; 2000US-0245291P.

FR 02-NOV-2000; 2000US-0245317P.

PR 07-NOV-2000; 2000US-0246562P.

PR 08-NOV-2000; 2000US-0246871P.

PR 26-JAN-2001; 2001US-0264389P.

PR 26-JAN-2001; 2001US-0264423P.

PR 29-JAN-2001; 2001US-0264799P.

XX (CURA-) CURAGEN CORP.

PI Grasse WM, Macdougall JR, Smithson G, Millet I, Stone DJ;  
PI Gunther E, Ellerman K, Alsobrook JP, Lepley DM, Burgess CE;  
PI Spytek KA, Edinger SR, Gangolli EA, Gorman L, Taupier RJ, Li L;  
PI Guo X, Fernandes ER, Vernet CAM, Tchernev VT, Casman SJ, Shenoy S;  
PI Mishra V, Furtak K, Baumgartner JC, Colman SD;  
XX WPI, 2003-140359/13.

DR N-PSDB; ACF03558.

XX New NOVX polypeptide useful for preventing or treating NOVX-associated

PT disorders, e.g. cancer, cardiomyopathy, atherosclerosis or diabetes, and

PT in chromosome mapping, tissue typing or pharmacogenomics.

XX Claim 1; Page 69; 346pp; English.

XX ACF03547 to ACF03570 encode the human NOVX proteins (I) given in ABR57412  
CC to ABR57435. (I) have cytostatic, cardiac, antiinflammatory, nontropic,  
CC immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic,  
CC antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, virucide,  
CC antiarthritic, hepatotropic, neuroprotective, antibacterial, relaxant,  
CC antiparasitic, anticonvulsant, hypotensive, vasotropic, antiparkinsonian,  
CC vulnerary, angiogenic and antiangiogenic activities, and can be used in  
CC gene therapy and vaccines. The NOVX polypeptides and their antibodies can  
CC be used to determine the presence or absence of (I) in a sample. The NOVX  
CC polypeptides, polynucleotides encoding them, and antibodies against them,  
CC are useful in manufacturing a medicament for treating or preventing a  
CC syndrome associated with a NOVX-associated disorder such as hypertension,  
CC cardiomyopathy, atherosclerosis, cancer, diabetes, asthma, inflammation,  
CC autoimmune disorders, allergies, blood disorders, obesity, acquired  
CC immunodeficiency syndrome (AIDS), immunoglobulin (Ig)A nephropathy,  
CC cirrhosis, arthritis, Alzheimer's disease, Parkinson's disease, goitre,  
CC infections (e.g. bacterial, viral, parasitic), stroke, muscular  
CC dystrophy, epilepsy, and other wasting disorders associated with chronic  
CC diseases. ACF03571 to ACF03644 represent PCR primers and probes for NOVX  
CC sequence, which are used in an example from the present invention

XX Sequence 149 AA;

Query Match 100.0%; Score 58; DB 6; Length 149;  
Best Local Similarity 100.0%; Pred. No. 0.31; Mismatches 0; Gaps 0;  
Matches 10; Conservative 0;

QY 1 RGRPGQPQQ 10

DB 105 RGRPGQPQQ 114

RESULT 10

ABR56769

ID ABR56769 standard; protein; 154 AA.

XX ABR56769;

AC ABR56769;

DT 30-JUL-2003 (first entry)

DE Human secreted protein SECP-44 SEQ ID NO:44.

XX Human; secreted protein; SECP; cytostatic; antiarteriosclerotic;

KW anticonvulsant; nontropic; neuroprotective; cerebroprotective; anti-HIV;

KW antiallergic; antiinflammatory; thyromimetic; gene therapy; cancer;

KW cell proliferative disorder; atherosclerosis; neurological disorder;

KW epilepsy; Huntington's disease; stroke; immune disorder; allergy; AIDS;

KW inflammatory disorder; developmental disorder; hypothyroidism;

XX Cushing's syndrome; infection.

OS Homo sapiens.

XX WO2003016506-A2.

PN 27-FEB-2003.

PD 15-AUG-2002; 2002WO-US027143.

XX

PR 17-AUG-2001; 2001US-0313249P.  
PR 24-AUG-2001; 2001US-0314752P.  
PR 07-SEP-2001; 2001US-0317818P.  
PR 07-SEP-2001; 2001US-0317824P.  
PR 21-SEP-2001; 2001US-0324040P.  
PR 24-SEP-2001; 2001US-0324586P.  
PR 02-NOV-2001; 2001US-0343980P.  
PR 28-NOV-2001; 2001US-0334229P.  
PR 13-FEB-2002; 2002US-0357002P.  
PR 06-MAR-2002; 2002US-0362439P.  
PR 19-MAR-2002; 2002US-0366041P.  
PR 30-APR-2002; 2002US-0376989P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Tang TY, Warren BA, Gietzen KJ, Lal PG, Yue H, Honchell CD;  
PI Lehr-Mason PM, Burford N, Xu Y, Baughn MR, Duggan BW, Tran UK;  
PI Lee EA, Forsythe IJ, Richardson TW, Lee S, Thangavelu K, Yue H;  
PI Emerling BM, Wallia NK, Azimzai Y, Sanjanwala B, Hafalia AJA;  
PI Borowsky ML, Nguyen DB, Ison CH, Astromoff A, Ding L, Lee SY;  
PI Becha SD, Ramkumar J, Gandhi AR, Jin P, Fu GK, Swarnakar A;  
XX WPI; 2003-278569/27.  
DR N-PSDB; ACC79069.  
XX New human secreted proteins (SECP), useful for diagnosing, treating and  
PT preventing diseases or conditions associated with the aberrant SECP  
PT expression e.g. cancer, AIDS, atherosclerosis, Huntington's disease,  
PT stroke, infections.  
XX Claim 1; Page 222; 286pp; English.  
XX ACC75026 to ACC79105 encode the human secreted proteins (I) given in  
CC ABR56726 to ABR56805, designated SECP-1 to SECP-80. SECP sequences can  
CC have cytostatic, antiarteriosclerotic, anticonvulsant, antiinflammatory,  
CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic and  
CC thymimetic activities, and can be used in gene therapy. The SECP  
CC proteins and polynucleotides can be used in diagnosing, treating and  
CC preventing diseases or conditions associated with the decreased  
CC expression or overexpression of SECP, such as cell proliferative (e.g.  
CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's  
CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and  
CC developmental (e.g. hypothyroidism, Cushing's syndrome) disorders, or  
CC infections. They are also useful in assessing the effects of exogenous  
CC compounds on the expression of nucleic acid and amino acid sequences of  
CC SECP. The SECP or its fragments are useful in screening compounds for  
CC effectiveness as agonist or antagonist of the polypeptides, or in  
CC altering the expression of the target polynucleotide and compounds that  
CC specifically bind to or modulate the activity of the polypeptide  
XX SQ Sequence 154 AA;  
Query Match 100.0%; Score 58; DB 6; Length 154;  
Best Local Similarity 100.0%; Pred. No. 0.32;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRPGQPPQQ 10  
Db 110 RGRPGQPPQQ 119  
RESULT 11  
ADA83798  
ID ADA83798 standard; protein; 166 AA.  
XX ADA83798;  
XX 20-NOV-2003 (first entry)  
XX Human PRH2 protein.  
DE human; marker; expressed sequence tag; EST; arabidopsis; tumour;  
XX stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;  
KW

XX vaccine.  
XX Homo sapiens.  
XX WO2002103028-A2.  
XX 27-DEC-2002.  
XX 30-MAY-2002; 2002WO-IB004189.  
XX 30-MAY-2001; 2001US-0293999P.  
PR 22-OCT-2001; 2001US-0330457P.  
PR 19-FEB-2002; 2002US-0357144P.  
XX (BIOM-) BIOMEDICAL CENT.  
XX Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;  
XX WPI; 2003-175241/17.  
DR N-PSDB; ADA83797.  
XX Determining if a nucleic acid is a marker for a phenotype/cell type of  
PT interest, by global comparison of expressed sequence tags known to be  
PT expressed in the phenotype/cell type with all ESTs expressed in normal  
PT tissue.  
XX Claim 29; Page 191-192; 516pp; English.  
XX The invention relates to a novel method for determining if a nucleic acid  
CC is a marker for a predetermined phenotype/cell type of interest from a  
CC biological species. The method comprises performing a global comparison  
CC of a group of expressed sequence tags (ESTs) known to be expressed in the  
CC phenotype/cell type of interest with all ESTs expressed in normal tissue  
CC in order to identify ESTs that are preferentially expressed in the  
CC phenotype/cell type of interest. A method of the invention is useful for  
CC determining whether a nucleic acid is a marker for a predetermined  
CC phenotype or cell type of interest from a biological species, preferably  
CC Arabidopsis or human. The cell type of interest is an abnormal cell such  
CC as a tumour cell, and the predetermined phenotype is a stress-induced  
CC phenotype such as hyperosmotic stress or high salt conditions. A method  
CC of the invention is also useful for determining the progression of colon  
CC cancer in a human, for detecting a tumour cell, and for regulating or  
CC preventing the growth of a tumour cell. An antibody of the invention is  
CC useful for detecting the absence or presence of peptides encoded by  
CC tumour-associated markers. A polypeptide of the invention is useful as an  
CC immunogen for vaccinating an animal. The present sequence represents a  
CC tumour-associated antigen of the invention.  
XX SQ Sequence 166 AA;  
Query Match 100.0%; Score 58; DB 6; Length 166;  
Best Local Similarity 100.0%; Pred. No. 0.34;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRPGQPPQQ 10  
Db 122 RGRPGQPPQQ 131  
RESULT 12  
ADC98216  
ID ADC98216 standard; protein; 166 AA.  
XX ADC98216;  
XX 01-JAN-2004 (first entry)  
XX Human salivary acidic proline-rich phosphoprotein (PRP).  
DE Human; salivary acidic proline-rich phosphoprotein; PRP; autoantigen;  
XX atopic dermatitis-inducing protein; salivary gland; IGE autoantibody;  
KW immunoglobulin E; mast cell activation; basophil activation; diagnosis;  
KW risk assessment; sensitisation remedy; dermatological; antiallergic;  
KW



KW antiinflammatory.  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO2003084991-A1.  
 XX  
 PD 16-OCT-2003.  
 XX  
 XX 04-APR-2003; 2003WO-JP004325.  
 PF  
 XX 08-APR-2002; 2002JP-00105425.  
 PR  
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 PA  
 XX Hide M, Yamamoto S, Tanaka T, Koro O;  
 PI  
 XX WPI: 2003-833567/77.  
 DR N-PSDB; ADC98215.  
 DR  
 XX  
 XX Atopic dermatitis-inducing proteins, applicable in diagnosis of including  
 PT risk of onset, and in developing sensitization remedies for the disease.  
 PT  
 XX  
 PS Claim 4; SEQ ID NO 2; 43pp; Japanese.  
 XX  
 XX The invention relates to the human atopic dermatitis-inducing proteins,  
 CC salivary acidic proline-rich phosphoprotein (PRP; ADC98216) and prolactin  
 CC -inducible protein (PIP; ADC98218), and their post-translationally  
 CC modified forms. These proteins are secreted by salivary or sweat glands  
 CC and bind to IGE autoantibodies, thereby activating mast cells and  
 CC basophils. The invention also relates to antigenic peptide fragments of  
 CC PRP or PIP, antibodies which bind to PRP or PIP, methods for diagnosing  
 CC atopic dermatitis or for determining whether an individual is at risk of  
 CC developing atopic dermatitis by determining the presence of PRP- or PIP-  
 CC specific antibodies or immune complexes, or by quantifying histamine  
 CC release, and sensitization remedies for atopic dermatitis containing PRP  
 CC and/or PIP or their peptides as the active ingredient. PRP, PIP and their  
 CC antibodies are useful in diagnosing atopic dermatitis, or for determining  
 CC whether an individual is at risk of developing atopic dermatitis. They  
 CC are also useful in developing sensitization remedies for the treatment of  
 CC atopic dermatitis. The present sequence represents the specifically  
 CC claimed human salivary acidic proline-rich phosphoprotein (PRP).  
 XX  
 SQ Sequence 166 AA;  
 Query Match 100.0%; Score 58; DB 7; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.34;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPGPPQQ 10  
 DB 122 RGRPGPPQQ 131  
 RESULT 13  
 AAB48775  
 ID AAB48775 standard; peptide; 9 AA.  
 XX  
 AC AAB48775;  
 XX  
 XX 09-MAR-2001 (first entry)  
 DT  
 XX Human saliva PRP-1 fragment (residues 106-114), SEQ ID NO:5.  
 DE  
 XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
 KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200069890-A1.  
 PN  
 XX 23-NOV-2000.  
 PD  
 XX

PF 11-MAY-2000; 2000WO-SE000930.  
 XX  
 PR 17-MAY-1999; 99SE-00001773.  
 XX  
 PA (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX  
 PI Stroemberg N, Johansson I;  
 XX  
 DR WPI: 2001-031923/04.  
 XX  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 PT  
 XX Claim 4; Page 24; 36pp; English.  
 PS  
 XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by actinomyces and streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX  
 SQ Sequence 9 AA;  
 Query Match 91.4%; Score 53; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRPGPPQQ 9  
 DB 1 RGRPGPPQQ 9  
 RESULT 14  
 AAW03557  
 ID AAW03557 standard; peptide; 44 AA.  
 XX  
 AC AAW03557;  
 XX  
 DT 10-DEC-1996 (first entry)  
 DE  
 XX Human proline-rich saliva glycoprotein P-C.  
 XX  
 KW Human; saliva; proline-rich glycoprotein; insulin; secretion; glucagon;  
 KW perfusion pancreas model; Wistar rat; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 XX JP08092281-A.  
 PN  
 XX 09-APR-1996.  
 PD  
 XX 27-SEP-1994; 94JP-00257540.  
 PF  
 XX 27-SEP-1994; 94JP-00257540.  
 PR  
 XX (TEIK ) TEIKOKU HORMONE MFG CO LTD.  
 PA  
 XX WPI: 1996-236094/24.  
 DR  
 XX New octadecapeptide for treating diabetes - acts as insulin secretion  
 PT promoter and glucagon secretion inhibitor.  
 XX  
 PS Disclosure; Page 2; 7pp; Japanese.  
 XX

CC This is the sequence of the human proline-rich saliva glycoprotein P-C  
 CC which was used to synthesise an octadecapeptide (AAW03556) which has an  
 CC insulin secretion promoting activity and a glucagon secretion inhibiting  
 CC activity. The peptide was generated by usual peptide synthesis method and  
 CC opt. contains Boc protective gps. at the N-terminal His and on the  
 CC residue during synthesis. These are removed by cleavage with HCl for 10  
 CC mins on ice. The insulin secretion promoting activity and glucagon  
 CC secretion inhibiting activity were measured in a perfusion pancreas model  
 CC in a male Wistar rat. Insulin secretion was 1.78 times than normal and  
 CC glucagon secretion was 0.77 times normal levels. The peptide can thus be  
 CC used to treat diabetes  
 XX  
 XX SQ Sequence 44 AA;

Query Match 91.4%; Score 53; DB 2; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 0.57;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGPPQQ 10  
 |||||  
 Db 1 GRPQGPPQQ 9

RESULT 15  
 AAP93320  
 ID AAP93320 standard; protein; 45 AA.

XX AC AAP93320;  
 XX DT 19-JUN-1990 (first entry)  
 XX DE P-C gene.  
 XX KW P-C gene; fused protein; silica gel absorbent.

XX OS Unidentified.

XX PN JP01051097-A.

XX PD 27-FEB-1989.

XX PF 21-AUG-1987; 87JP-00208616.

XX PR 11-DEC-1979; 79JP-00160528.

XX PA (TORA ) TORAY IND INC.

XX DR WPI; 1989-104005/14.

XX DR N-PSDB; AAN92373.

XX PT Purificn. of fused protein(s) - comprises contacting fused protein with  
 PT silica gel absorbent and eluting.

XX PS Fig 1; page 6; 8pp; Japanese.

XX CC The DNA encoding it is prepd. as 14 oligomers. The 5' terminal of 12 of  
 CC the oligomers is phosphorylated, complementary 2 or 3 oligomers are  
 CC annealed., and the resulting 6 ds DNAs are combined to form the P-C gene  
 XX

SQ Sequence 45 AA;

Query Match 91.4%; Score 53; DB 1; Length 45;  
 Best Local Similarity 100.0%; Pred. No. 0.58;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRPQGPPQQ 10  
 |||||  
 Db 2 GRPQGPPQQ 10

Search completed: April 6, 2004, 16:06:38  
 Job time : 55.9813 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 39.2523 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-6  
Perfect score: 58  
Sequence: 1 RGRPQPPQ 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

```

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCT05_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	100.0	132	9	US-09-864-761-43644 Sequence 43644, A
2	58	100.0	166	14	US-10-157-031-80 Sequence 80, Appl
3	47	81.0	149	12	US-10-424-599-171613 Sequence 171613,
4	43	74.1	803	14	US-10-156-761-12958 Sequence 12958, A
5	42	72.4	159	12	US-10-424-599-212956 Sequence 212956,
6	42	72.4	224	12	US-10-425-114-62993 Sequence 62993, A
7	41	70.7	130	12	US-10-424-599-215801 Sequence 215801,
8	41	70.7	807	14	US-10-112-372-11 Sequence 11, Appl
9	40	69.0	92	14	US-10-029-386-23755 Sequence 23755, A
10	40	69.0	150	12	US-10-424-599-146985 Sequence 146985,
11	40	69.0	274	9	US-09-850-887-4 Sequence 4, Appli
12	40	69.0	458	15	US-10-369-493-2375 Sequence 2375, Ap
13	40	69.0	842	12	US-10-112-944-373 Sequence 373, Appl
14	40	69.0	19723	15	US-10-084-846A-5 Sequence 5, Appl1
15	39	67.2	37	12	US-10-424-599-227554 Sequence 227554,

16	39	67.2	199	12	US-10-425-114-60578 Sequence 60578, A
17	39	67.2	209	15	US-10-108-260A-4370 Sequence 4370, Ap
18	39	67.2	268	12	US-10-425-114-59678 Sequence 59678, A
19	39	67.2	292	12	US-10-425-114-52999 Sequence 52999, A
20	39	67.2	300	12	US-10-425-114-64837 Sequence 64837, A
21	39	67.2	311	12	US-10-425-114-45262 Sequence 45262, A
22	39	67.2	311	12	US-10-425-114-59180 Sequence 59180, A
23	39	67.2	311	12	US-10-425-114-61343 Sequence 61343, A
24	39	67.2	339	12	US-10-425-114-45434 Sequence 45434, A
25	39	67.2	451	15	US-10-286-829-132 Sequence 132, App
26	39	67.2	406	12	US-10-425-114-53918 Sequence 53918, A
27	39	67.2	2906	15	US-10-015-115-60 Sequence 60, Appl
28	39	67.2	2911	15	US-10-295-027-162 Sequence 162, App
29	38	65.5	52	12	US-10-424-599-277859 Sequence 277859,
30	38	65.5	76	12	US-10-424-599-204459 Sequence 204459,
31	38	65.5	110	12	US-10-424-599-193123 Sequence 193123,
32	38	65.5	116	12	US-10-424-599-280434 Sequence 280434,
33	38	65.5	117	12	US-10-424-599-162502 Sequence 162502,
34	38	65.5	121	12	US-10-424-599-271545 Sequence 271545,
35	38	65.5	162	12	US-10-424-599-219187 Sequence 219187,
36	38	65.5	178	14	US-10-029-137-2 Sequence 2, Appl1
37	38	65.5	178	14	US-10-029-137-6 Sequence 6, Appli
38	38	65.5	179	12	US-10-424-599-159940 Sequence 159940,
39	38	65.5	242	9	US-09-861-097-23 Sequence 23, Appl
40	38	65.5	242	10	US-09-861-098-23 Sequence 23, Appl
41	38	65.5	251	15	US-10-369-493-16046 Sequence 16046, A
42	38	65.5	432	15	US-10-108-260A-3406 Sequence 3406, Ap
43	38	65.5	432	15	US-10-369-493-20988 Sequence 20988, A
44	38	65.5	496	15	US-10-369-493-20988 Sequence 20988, A
45	38	65.5	518	14	US-10-364-360-23 Sequence 23, Appl

## ALIGNMENTS

### RESULT 1

```

US-09-864-761-43644
; Sequence 43644, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30

```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43644
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
; OTHER INFORMATION: EST_HUMAN HIT: BF089785.1, EVALUATE 1.00e-01
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUATE 7.40e-02
; US-09-864-761-43644

Query Match      100.0%; Score 58; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPGQPQQ 10
Db 88 RGRPGQPQQ 97

RESULT 2
US-10-157-031-80
; Sequence 80, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-80

Query Match      100.0%; Score 58; DB 14; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRPGQPQQ 10
Db 122 RGRPGQPQQ 131

RESULT 3
US-10-424-599-171613
; Sequence 171613, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 171613
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_125982C.1.pap
US-10-424-599-171613

Query Match      81.0%; Score 47; DB 12; Length 149;
Best Local Similarity 80.0%; Pred. No. 28;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RGRPGQPQQ 10
Db 55 RGRPGQPQQ 64

RESULT 4
US-10-156-761-12958
; Sequence 12958, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12958
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12958

Query Match      74.1%; Score 43; DB 14; Length 803;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GRPQGP 8
Db 195 GRPQGP 201

RESULT 5
US-10-424-599-212996
; Sequence 212996, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
```

```
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 212996
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(159)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_34360C.1.pap
US-10-424-599-212996

Query Match          72.4%; Score 42; DB 12; Length 159;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      2 GRPQGPPQQ 10
DB      147 GRPEGPPRK 155

RESULT 6
US-10-425-114-62999
; Sequence 62999, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62999
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3115-038-D1_PLI.pap
US-10-425-114-62999

Query Match          72.4%; Score 42; DB 12; Length 224;
Best Local Similarity 70.0%; Pred. No. 2e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 RGRPQGPQQ 10
DB      155 RGAPAGPPQR 164

RESULT 7
US-10-424-599-215801
; Sequence 215801, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 215801
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3689C.1.pap
US-10-424-599-215801

Query Match          70.7%; Score 41; DB 12; Length 130;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GRPQGPPQQ 9
DB      64 GPGQPPQQ 71

RESULT 8
US-10-112-372-11
; Sequence 11, Application US/10112372
; Publication No. US20030186249A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human TARPP Genes and Polypeptides
; FILE REFERENCE: 16U 105 R1
; CURRENT FILING DATE: 2002-04-01
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Mus musculus
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3689C.1.pap
US-10-112-372-11

Query Match          70.7%; Score 41; DB 14; Length 807;
Best Local Similarity 77.8%; Pred. No. 8.7e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 GRPQGPPQQ 10
DB      511 GPGQPPQQ 519

RESULT 9
US-10-029-386-29755
; Sequence 29755, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AECOMICA-X-2
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29755
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR17.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.71
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.61
```

```
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 215801
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3689C.1.pap
US-10-424-599-215801

Query Match          70.7%; Score 41; DB 12; Length 130;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GRPQGPPQQ 9
DB      64 GPGQPPQQ 71

RESULT 8
US-10-112-372-11
; Sequence 11, Application US/10112372
; Publication No. US20030186249A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human TARPP Genes and Polypeptides
; FILE REFERENCE: 16U 105 R1
; CURRENT FILING DATE: 2002-04-01
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Mus musculus
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3689C.1.pap
US-10-112-372-11

Query Match          70.7%; Score 41; DB 14; Length 807;
Best Local Similarity 77.8%; Pred. No. 8.7e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 GRPQGPPQQ 10
DB      511 GPGQPPQQ 519

RESULT 9
US-10-029-386-29755
; Sequence 29755, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AECOMICA-X-2
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29755
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR17.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.71
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.61
```

```
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.99
; OTHER INFORMATION: SWISSPROT HIT: Q35569, EVALUE 1.10e-02
US-10-029-386-29755

Query Match          69.0%; Score 40; DB 14; Length 92;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 GRPQGPQQ 9
      |||:||||:
Db      47 GRPGRPK 54

RESULT 10
US-10-424-599-146985
; Sequence 146985, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 146985
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_103748C.1.pep
US-10-424-599-146985

Query Match          69.0%; Score 40; DB 12; Length 150;
Best Local Similarity 60.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 RGPQGPQQ 10
      |||:||||:
Db      104 RGRPGGPKK 113

RESULT 11
US-09-850-887-4
; Sequence 4, Application US/09850887
; Patent No. US20020009778A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina
; TITLE OF INVENTION: THYROID AND PITUITARY MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/850,887
; FILING DATE: 07-May-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/087,678
```

```
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0535 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: g206712
; SEQUENCE DESCRIPTION: SEQ ID NO: 4 :
US-09-850-887-4

Query Match          69.0%; Score 40; DB 9; Length 274;
Best Local Similarity 87.5%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GRPQGPQQ 9
      |||:||||:
Db      263 GRPQGPQQ 270

RESULT 12
US-10-369-493-2375
; Sequence 2375, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2375
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(458)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-2375

Query Match          69.0%; Score 40; DB 15; Length 458;
Best Local Similarity 87.5%; Pred. No. 7.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GRPQGPQQ 9
      |||:||||:
Db      395 GRPQGPQQ 402

RESULT 13
US-10-112-944-373
; Sequence 373, Application US/10112944
; Publication No. US20040048249A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
```

APPLICANT: Weng, Gezhi  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Xue, Aidong J.  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Wehrman, Tom  
APPLICANT: Ghosh, Malabika  
APPLICANT: Wang, Dunrui  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wang, Zhiwei  
TITLE OF INVENTION: No. US20040048249A1el Nucleic Acids and  
TITLE OF INVENTION: Secreted Polypeptides  
FILE REFERENCE: 805A  
CURRENT APPLICATION NUMBER: US/10/112,944  
CURRENT FILING DATE: 2002-03-28  
PRIOR APPLICATION NUMBER: US 09/488,725  
PRIOR FILING DATE: 2000-01-21  
PRIOR APPLICATION NUMBER: US 09/491,404  
PRIOR FILING DATE: 2000-01-25  
PRIOR APPLICATION NUMBER: US 09/496,914  
PRIOR FILING DATE: 2000-02-03  
PRIOR APPLICATION NUMBER: US 09/515,126  
PRIOR FILING DATE: 2000-02-28  
PRIOR APPLICATION NUMBER: US 09/519,705  
PRIOR FILING DATE: 2000-03-07  
PRIOR APPLICATION NUMBER: US 09/540,217  
PRIOR FILING DATE: 2000-03-31  
PRIOR APPLICATION NUMBER: US 09/552,929  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: US 09/577,408  
PRIOR FILING DATE: 2000-05-18  
NUMBER OF SEQ ID NOS: 924  
SOFTWARE: PC-FL-Genes Version 5.0  
SEQ ID NO 373  
LENGTH: 842  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-112-944-373

Query Match 69.0%; Score 40; DB 12; Length 842;  
Best Local Similarity 66.7%; Pred. No. 1.2e+03;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPPQ 10  
DB 456 GRPMGPPR 454

RESULT 14  
US-10-084-846A-5  
Sequence 5, Application US/10084846A  
Publication No. US20040006026A1  
GENERAL INFORMATION:  
APPLICANT: WEITNAUER, GABRIELE  
APPLICANT: MOHLENWEG, AGNES  
APPLICANT: TREFFZER, AXEL  
APPLICANT: BECHTHOLD, ANDREAS  
TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
FILE REFERENCE: 1974-005  
CURRENT APPLICATION NUMBER: US/10/084,846A  
CURRENT FILING DATE: 2003-02-25  
PRIOR APPLICATION NUMBER: PCT/EP01/09815  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: DE 101 09 166.4  
PRIOR FILING DATE: 2001-02-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 3.2  
SEQ ID NO 5  
LENGTH: 19723  
TYPE: PRT  
ORGANISM: Streptomyces viridochromogenes  
FEATURE:  
OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.

OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.  
US-10-084-846A-5

Query Match 69.0%; Score 40; DB 15; Length 19723;  
Best Local Similarity 77.8%; Pred. No. 2.1e+04;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGPPQ 9  
DB 10188 RGRPGPQ 10196

## RESULT 15

US-10-424-599-227554  
Sequence 227554, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 227554  
LENGTH: 37  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_4750C.1.pep  
US-10-424-599-227554

Query Match 67.2%; Score 39; DB 12; Length 37;  
Best Local Similarity 66.7%; Pred. No. 1e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPPQ 10  
DB 12 GAPKGPQK 20

Search completed: April 6, 2004, 17:05:54  
Job time : 40.2523 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 14.6729 seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-6  
Perfect score: 58  
Sequence: 1 RGRPGPPQ 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pap.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pap.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pap.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pap.\*  
5: /cgn2\_6/ptodata/2/iaa/pCTUS\_COMB.pap.\*  
6: /cgn2\_6/ptodata/2/iaa/backfile1.pap.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	79.3	423	4	US-09-489-039A-7834
2	45	77.6	391	4	US-09-252-991A-20316
3	44	75.9	82	2	US-08-925-237-4
4	44	75.9	124	2	US-08-925-237-2
5	43	74.1	366	4	US-09-252-991A-17018
6	42	72.4	710	4	US-09-252-991A-22478
7	41	70.7	27	3	US-08-658-136-29
8	41	70.7	660	4	US-09-252-991A-22842
9	41	70.7	725	4	US-09-252-991A-23752
10	41	70.7	1614	4	US-09-052-469-2
11	41	70.7	1614	4	US-08-422-582-2
12	41	70.7	1614	4	US-09-052-262-2
13	41	70.7	4339	4	US-09-052-469-6
14	41	70.7	4339	4	US-08-422-582-6
15	41	70.7	4339	4	US-09-052-262-6
16	40	69.0	238	4	US-09-252-991A-23387
17	40	69.0	570	4	US-09-252-991A-21238
18	40	69.0	977	4	US-09-252-991A-16655
19	39	67.2	160	4	US-09-252-991A-27129
20	39	67.2	195	4	US-09-252-991A-22884
21	39	67.2	456	4	US-09-328-352-6946
22	39	67.2	478	3	US-08-155-888-2
23	38	65.5	151	4	US-09-621-976-4549
24	38	65.5	242	4	US-08-220-602B-23
25	38	65.5	318	4	US-09-452-991A-20530
26	38	65.5	516	4	US-09-625-972-23
27	38	65.5	543	2	US-08-469-412A-7

28	38	65.5	543	3	US-09-021-715-7	Sequence 7, Appli
29	38	65.5	878	4	US-09-252-991A-31974	Sequence 31974, A
30	37	63.8	16	3	US-08-602-999A-191	Sequence 191, App
31	37	63.8	16	4	US-09-500-124-191	Sequence 191, App
32	37	63.8	30	4	US-08-948-131-2	Sequence 2, Appli
33	37	63.8	30	4	US-08-948-131-4	Sequence 4, Appli
34	37	63.8	87	4	US-09-621-976-7698	Sequence 7698, Ap
35	37	63.8	135	4	US-09-252-991A-21465	Sequence 21466, A
36	37	63.8	219	4	US-09-252-991A-17169	Sequence 17169, A
37	37	63.8	262	4	US-09-252-991A-21401	Sequence 21401, A
38	37	63.8	276	4	US-09-489-039A-9778	Sequence 9778, Ap
39	37	63.8	291	4	US-09-252-991A-19638	Sequence 19638, A
40	37	63.8	311	3	US-08-911-423-8	Sequence 8, Appli
41	37	63.8	323	1	US-08-469-667-16	Sequence 16, Appl
42	37	63.8	323	3	US-08-946-914-2	Sequence 2, Appli
43	37	63.8	323	4	US-09-224-110-16	Sequence 16, Appli
44	37	63.8	323	4	US-09-656-450-2	Sequence 2, Appli
45	37	63.8	323	5	PCT-US95-07289-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1

US-09-489-039A-7834  
; Sequence 7834, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 7834  
; LENGTH: 423  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-7834

Query Match

Best Local Similarity 79.3%; Score 46; DB 4; Length 423;  
Matches 7; Conservative 77.8%; Pred. No. 16;  
Mismatches 2; Indels 0; Gaps 0;

Qy 1 RGRPGPPQ 9  
Db 194 KGRPGPPR 202

RESULT 2

US-09-252-991A-20316  
; Sequence 20316, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20316  
; LENGTH: 391  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20316



```
Query Match 77.6%; Score 45; DB 4; Length 391;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGQPPQ 10
DB 105 RGRPAGPFR 114

RESULT 3
US-08-925-237-4
; Sequence 4, Application US/08925237
; Patent No. 5981720
; GENERAL INFORMATION:
; APPLICANT: Azen, Edwin A.
; APPLICANT: Pan, David
; TITLE OF INVENTION: Human Salivary Proteins And Fragments
; TITLE OF INVENTION: Thereof Having Alpha-Glucosidase Inhibitory Activity
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kent Barta
; STREET: 100 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: USA
; ZIP: 53202-4108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,237
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Barta, Kent S.
; REGISTRATION NUMBER: 29,042
; REFERENCE/DOCKET NUMBER: 96429/9002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 257-3501
; TELEFAX: (608) 283-2275
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 82 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-925-237-4

Query Match 75.9%; Score 44; DB 2; Length 82;
Best Local Similarity 77.8%; Pred. No. 7.1;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10
DB 4 GKPGGPFF 12

RESULT 4
US-08-925-237-2
; Sequence 2, Application US/08925237
; Patent No. 5981720
; GENERAL INFORMATION:
; APPLICANT: Azen, Edwin A.
; APPLICANT: Pan, David
; TITLE OF INVENTION: Human Salivary Proteins And Fragments
; TITLE OF INVENTION: Thereof Having Alpha-Glucosidase Inhibitory Activity
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kent Barta
; STREET: 100 East Wisconsin Avenue
```

```
; CITY: Milwaukee
; STATE: WI
; COUNTRY: USA
; ZIP: 53202-4108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,237
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Barta, Kent S.
; REGISTRATION NUMBER: 29,042
; REFERENCE/DOCKET NUMBER: 96429/9002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 257-3501
; TELEFAX: (608) 283-2275
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-925-237-2

Query Match 75.9%; Score 44; DB 2; Length 124;
Best Local Similarity 77.8%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10
DB 4 GRPQGPQQ 12

RESULT 5
US-09-252-991A-17018
; Sequence 17018, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17018
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-17018

Query Match 74.1%; Score 43; DB 4; Length 366;
Best Local Similarity 70.0%; Pred. No. 39;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGRPGPQQ 10
DB 135 RGRPAGPFR 144

RESULT 6
US-09-252-991A-22478
; Sequence 22478, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
```

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 22478  
LENGTH: 710

TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22478

Query Match 72.4%; Score 42; DB 4; Length 710;  
Best Local Similarity 70.0%; Pred. No. 98;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RGRPGGPPQQ 10  
Db 605 RGRPRGRPEQ 614

RESULT 7  
US-08-658-136-29  
Sequence 29, Application US/08658136  
Patent No. 6071717  
GENERAL INFORMATION:  
APPLICANT: KLINGER, KATHERINE W  
APPLICANT: LANDES, GREGORY M  
APPLICANT: BURN, TIMOTHY C  
APPLICANT: CONNORS, TIMOTHY D  
APPLICANT: DACKOWSKI, WILLIAM  
APPLICANT: GERMINO, GREGORY  
APPLICANT: QIAN, FENG  
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENZYME CORPORATION  
STREET: ONE MOUNTAIN ROAD  
CITY: FRAMINGHAM  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 01701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/658,136  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: LASSEN, ELIZABETH  
REGISTRATION NUMBER: 31,845  
REFERENCE/DOCKET NUMBER: GEN4-17.8  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 508-872-8400  
TELEFAX: 508-872-5415  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-658-136-29

Query Match 70.7%; Score 41; DB 3; Length 27;  
Best Local Similarity 77.8%; Pred. No. 7;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 GRPQGGPPQQ 10  
Db 11 GRPQGGPPQQ 19

RESULT 8  
US-09-252-991A-22842  
Sequence 22842, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 22842  
LENGTH: 660  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22842

Query Match 70.7%; Score 41; DB 4; Length 660;  
Best Local Similarity 70.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RGRPGGPPQQ 10  
Db 577 RGRPGDPPEQ 586

RESULT 9  
US-09-252-991A-23752  
Sequence 23752, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 23752  
LENGTH: 725  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23752

Query Match 70.7%; Score 41; DB 4; Length 725;  
Best Local Similarity 70.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RGRPGGPPQQ 10  
Db 621 RGRPGDPPEQ 630

RESULT 10  
US-09-052-469-2  
Sequence 2, Application US/09052469  
Patent No. 6380360

```

; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,469
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-April-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74165
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1614 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
;
US-09-052-469-2
Query Match 70.7%; Score 41; DB 4; Length 1614;
Best Local Similarity 77.8%; Pred. No. 2.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10
Db 1524 GREQGPQQ 1532

RESULT 11
US-08-422-582-2
; Sequence 2, Application US/08422582
; Patent No. 6485960
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 75 State Street
; CITY: Boston
; STATE: MA

```

```

; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,582
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/53313 (MRC-006xx)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1614 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
;
US-08-422-582-2
Query Match 70.7%; Score 41; DB 4; Length 1614;
Best Local Similarity 77.8%; Pred. No. 2.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10
Db 1524 GREQGPQQ 1532

RESULT 12
US-09-052-262-2
; Sequence 2, Application US/09052262
; Patent No. 6656681
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,262
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-April-1995

```

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9507766.5  
;; FILING DATE: 13-APR-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9411900.5  
;; FILING DATE: 14-JUN-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB PCT/GB94/02822  
;; FILING DATE: 23-DEC-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9326470.3  
;; FILING DATE: 24-DEC-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Williams, Ph.D., Kathleen M.  
;; REGISTRATION NUMBER: 34,380  
;; REFERENCE/DOCKET NUMBER: 3265/74118  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 345-9100  
;; TELEFAX: (617) 345-9111  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1614 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-052-469-2

Query Match 70.7%; Score 41; DB 4; Length 1614;  
Best Local Similarity 77.8%; Pred. No. 2.9e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GRQGGPPQQ 10  
|: |||||  
Db 1524 GQEQGPPQQ 1532

RESULT 13  
US-09-052-469-6  
; Sequence 6, Application US/09052469  
; Patent No. 6380360  
; GENERAL INFORMATION:  
; APPLICANT: Harris et al.  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: One Financial Center  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk, 3.50 inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/052,469  
; FILING DATE: Concurrently herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/422,582  
; FILING DATE: 14-April-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9507766.5  
; FILING DATE: 13-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9411900.5  
; FILING DATE: 14-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB PCT/GB94/02822  
; FILING DATE: 23-DEC-1994  
; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: GB 9326470.3  
;; FILING DATE: 24-DEC-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Williams, Ph.D., Kathleen M.  
;; REGISTRATION NUMBER: 34,380  
;; REFERENCE/DOCKET NUMBER: 3265/74165  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 345-9100  
;; TELEFAX: (617) 345-9111  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4339 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-052-469-6

Query Match 70.7%; Score 41; DB 4; Length 4339;  
Best Local Similarity 77.8%; Pred. No. 7.1e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GRQGGPPQQ 10  
|: |||||  
Db 4249 GQEQGPPQQ 4257

RESULT 14  
US-08-422-582-6  
; Sequence 6, Application US/08422582  
; Patent No. 6485960  
; GENERAL INFORMATION:  
; APPLICANT: Harris et al.  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 75 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk, 3.50 inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/422,582  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9411900.5  
; FILING DATE: 14-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB PCT/GB94/02822  
; FILING DATE: 23-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9507766.5  
; FILING DATE: 13-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9326470.3  
; FILING DATE: 24-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams, Ph.D., Kathleen M.  
; REGISTRATION NUMBER: 34,380  
; REFERENCE/DOCKET NUMBER: 3265/53313 (MRC-006xx)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 345-9100  
; TELEFAX: (617) 345-9111  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4339 amino acids  
; TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-422-582-6

Query Match 70.7%; Score 41; DB 4; Length 4339;  
Best Local Similarity 77.8%; Pred. No. 7.1e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10  
|: |||||

DB 4249 GQEQGPQQ 4257

## RESULT 15

US-09-052-262-6  
Sequence 6, Application US/09052262  
Patent No. 6856681  
GENERAL INFORMATION:  
APPLICANT: Harris et al.  
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Barner & Witcoff, Ltd.  
STREET: One Financial Center  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02111

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk, 3.50 inch  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/052.262

FILING DATE: Concurrently herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/422,582

FILING DATE: 14-April-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9507766.5

FILING DATE: 13-APR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9411900.5

FILING DATE: 14-JUN-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB PCT/GB94/02822

FILING DATE: 23-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9326470.3

FILING DATE: 24-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: Williams, Ph.D., Kathleen M.

REGISTRATION NUMBER: 34,380

REFERENCE/DOCKET NUMBER: 3265/74118

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 345-9100

TELEFAX: (617) 345-9111

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 4339 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-052-262-6

## Query Match

Best Local Similarity 70.7%; Score 41; DB 4; Length 4339;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GRPQGPQQ 10  
|: |||||

Db 4249 GQEQGPQQ 4257

Search completed: April 6, 2004, 16:19:40  
Job time : 15.6729 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 11.215 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-7

Perfect score: 62

Sequence: 1 GCHPRPRGR 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: piri:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	166	1	PIHUSC
2	62	100.0	166	2	salivary proline-r
3	62	100.0	171	2	salivary proline-r
4	49	79.0	340	2	proline-rich phosph
5	49	79.0	438	2	erythromycin resis
6	46	74.2	325	2	probable histidyl-
7	46	74.2	326	2	pseudouridylyate sy
8	46	74.2	326	2	ftsh suppressor pr
9	46	74.2	326	2	suppressor of fsh
10	46	74.2	326	2	suppressor of fsh
11	45	72.6	55	2	pseudouridylyate sy
12	44	71.0	561	2	DNA-binding protei
13	44	71.0	1852	2	hypothetical prote
14	43.5	70.2	707	2	calcium channel pr
15	43	69.4	521	1	PTB-associated spl
16	43	69.4	1596	2	glycoprotein gill
17	42	67.7	218	2	neurogenic locus m
18	42	67.7	448	2	hypothetical prote
19	42	67.7	453	2	alpha-2B-adrenergi
20	42	67.7	455	2	alpha-2B-adrenergi
21	42	67.7	479	2	alpha-C2 adrener
22	42	67.7	485	2	conserved hypothet
23	41	66.1	54	2	probable hydroxyla
24	41	66.1	54	2	hypothetical prote
25	41	66.1	414	2	hypothetical prote
26	41	66.1	460	2	zinc finger protei
27	41	66.1	481	1	penicillin-binding
28	41	66.1	550	2	trophoblast-endoth
29	41	66.1	924	2	initiation factor
			1466	2	CL1A4 protein - ra

## ALIGNMENTS

### RESULT 1

#### PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N/Alternate names: salivary acidic proline-rich protein PRH2

C/Contents: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotei

C/Species: Homo sapiens (man)

C/Date: 31-Mar-1981 #sequence revision 12-Apr-1996 #text change 08-Dec-2000

C/Accession: A25372; A19803; E57868; A92277; A92254; A91954; S02564; S02563; J

R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A/Title: Differential RNA splicing and post-translational cleavages in the human saliv

A/Reference number: A92492; MUID:85289325; PMID:2993301

A/Accession: A25372

A/Molecule type: mRNA

A/Residues: 1-166 <VAE>

A/Cross-references: GB:K03202; NID:gl90481; PIDN:AAA60183.1; PID:gl90482

R/Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A/Title: Primary structure of the active tryptic fragments of human and monkey salivar

A/Reference number: A91757; MUID:81191179; PMID:7228490

A/Accession: A19803

A/Molecule type: protein

A/Residues: 17-46 <SCH>

R/Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A/Title: Structures of two HaeIII-type genes in the human salivary proline-rich protei

A/Reference number: A57868; MUID:86196106; PMID:3009472

A/Accession: B57868

A/Molecule type: DNA

A/Residues: 1-166 <KIM>

A/Cross-references: GB:M13058; NID:gl90513; PIDN:AAA98808.1; PID:gl90514

R/Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A/Title: The primary structure of a salivary calcium-binding proline-rich phosphoprote

A/Reference number: A92277; MUID:80204368; PMID:7380845

A/Contents: protein C

A/Accession: A92277

A/Molecule type: protein

A/Residues: 17-19, 'N', 21-166 <WON>

A/Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite for

R/Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A/Title: The complete primary structure of a proline-rich phosphoprotein from human sa

A/Reference number: A92254; MUID:79173237; PMID:438215

A/Contents: protein A

A/Accession: A92254

A/Molecule type: protein

A/Residues: 17-19, 'N', 21-122 <W02>

R/Schlesinger, D.H.; Hay, D.I.

In Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gros

A/Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A/Reference number: A94425

latrophilin-1, bra  
CL1A4 protein - ra  
latrophilin-1, bra  
CL1A4 protein - ra  
CL1A4 protein - ra  
phosphoinositide-b  
hypothetical prote  
fatty-acid synthas  
hypothetical prote  
hypothetical prote  
cysteine proteinas  
oryzacystatin II -  
hypothetical prote  
hypothetical prote  
very hypothetical  
hypothetical prote

30 41 1467 2 T18411  
31 41 1471 2 T17149  
32 41 1472 2 T18413  
33 41 1510 2 T17145  
34 41 1515 2 T17156  
35 41 1544 2 E59431  
36 41 1544 2 T30637  
37 41 2133 2 JC4743  
38 41 2796 2 JC4743  
39 40 86 2 S00945  
40 64.5 100 2 C72667  
41 40 64.5 106 2 S13027  
42 40 64.5 107 2 A38375  
43 40 64.5 117 2 C72511  
44 40 64.5 129 2 F71241  
45 40 64.5 146 2 T38048  
46 40 64.5 185 2 F91020

A/Accession: A94425  
 A/Molecule type: protein  
 A/Residues: 17-122 <SC2>  
 A/Note: the authors call this protein PRP-4  
 R/Isemura, S.; Saichou, E.; Sanada, K.  
 J. Biochem. 87, 1071-1077, 1980  
 A/Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to the human salivary proline-rich proteins (PRP)  
 A/Reference number: A91954; MUID:80227634; PMID:7390979  
 A/Contents: peptide P-C  
 A/Accession: A91954  
 A/Molecule type: protein  
 A/Residues: 123-166 <ISE>  
 R/Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A/Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A/Reference number: S02562; MUID:89061650; PMID:3196309  
 A/Accession: S02564  
 A/Molecule type: protein  
 A/Residues: 17-166 <HAY>  
 A/Accession: S02563  
 A/Molecule type: protein  
 A/Residues: 47-71 <HA2>  
 R/Schlesinger, D.H.; Hay, D.I.  
 Int. J. Pept. Protein Res. 27, 373-379, 1986  
 A/Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of the covalent structure of the cov  
 A/Reference number: JP0106; MUID:8622916; PMID:3710693  
 A/Accession: JP0106  
 A/Molecule type: protein  
 A/Residues: 17-161, Q'163-166 <SC3>  
 A/Experimental source: parotid gland  
 R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure and possible origin of the non-glycosylated basic proline-rich proteins  
 A/Reference number: S06153; MUID:90088384; PMID:2688632  
 A/Accession: G38355  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 123-166 <KAU>  
 R/Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A/Title: Primary structure and possible origin of the non-glycosylated basic proline-rich proteins  
 A/Reference number: S06153; MUID:90088384; PMID:2688632  
 A/Accession: S06153  
 A/Molecule type: protein  
 A/Residues: 123-166 <ROB>  
 C/Comment: The proposed biological functions are a highly potent inhibitor of crystal growth  
 C/Genetics:  
 A/Gene: GDB:PRH2  
 A/Cross-references: GDB:119516; OMIM:168790  
 A/Map position: 12p13.2-12p13.2  
 A/Introns: 22/1; 34/1  
 C/Superfamily: proline-rich protein  
 C/Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva  
 F.1-16/Domain: signal sequence #status predicted <SIG>  
 F.17-166/Product: protein A #status experimental <PRC>  
 F.17-122/Product: protein C #status experimental <PRA>  
 F.17-46/Region: apatitic mineral binding  
 F.47-71/Product: PRP-3 #status experimental <PRP3>  
 F.123-166/Product: peptide P-C #status experimental <PPC>  
 F.17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
 F.24,38/Binding site: phosphate (Ser) (covalent) #status experimental  
 Query Match 100.0%; Score 62; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.021;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GGHPRPRGR 10  
 Db 115 GGHPRPRGR 124  
 RESULT 2  
 B25372  
 salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human  
 C/Species: Homo sapiens (man)  
 C/Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text change 20-Aug-1999  
 C/Accession: B25372; A57868; S02562; G38355; S06153; B27307  
 R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 11123-11130, 1985  
 A/Title: Differential RNA splicing and post-translational cleavages in the human salivary proline-rich phosphoprotein precursor PRH1 (allele PIF)  
 A/Reference number: A92492; MUID:85289325; PMID:2993301  
 A/Accession: B25372  
 A/Molecule type: mRNA  
 A/Residues: 1-166 <MAE>  
 A/Cross-references: GB:K03203; NID:G190483; PIDN:AAA60184.1; PID:G190484  
 R/Kim, H.S.; Maeda, N.  
 J. Biol. Chem. 261, 6712-6718, 1986  
 A/Title: Structures of two HaellI-type genes in the human salivary proline-rich protein PRP-1  
 A/Reference number: A57868; MUID:86196106; PMID:3009472  
 A/Accession: A57868  
 A/Molecule type: DNA  
 A/Residues: 1-166 <KIM>  
 A/Cross-references: GB:M13057; NID:G190511; PIDN:AAA98807.1; PID:G190512  
 R/Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A/Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A/Reference number: S02562; MUID:89061650; PMID:3196309  
 A/Accession: S02562  
 A/Molecule type: protein  
 A/Residues: 47-71 <HAY>  
 R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure and possible origin of the non-glycosylated basic proline-rich proteins  
 A/Reference number: S06153; MUID:90088384; PMID:2688632  
 A/Accession: S06153  
 A/Molecule type: protein  
 A/Residues: 123-166 <KAU>  
 R/Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A/Title: Primary structure and possible origin of the non-glycosylated basic proline-rich proteins  
 A/Reference number: S06153; MUID:90088384; PMID:2688632  
 A/Accession: S06153  
 A/Molecule type: protein  
 A/Residues: 123-166 <ROB>  
 R/Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A/Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein PRP-1  
 A/Reference number: A27307; MUID:88074309; PMID:3687941  
 A/Contents: allele Pa  
 A/Accession: B27307  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: DNA  
 A/Residues: 17-41, 'L', 43-118, 'C', 120-166 <AZE>  
 A/Cross-references: EMBL:K03203  
 C/Genetics:  
 A/Gene: GDB:PRH1  
 A/Cross-references: GDB:119515; OMIM:168730  
 A/Map position: 12p13.2-12p13.2  
 A/Introns: 22/1; 34/1  
 C/Superfamily: proline-rich protein  
 C/Keywords: phosphoprotein; saliva; tandem repeat  
 Query Match 100.0%; Score 62; DB 2; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.021;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GGHPRPRGR 10  
 Db 115 GGHPRPRGR 124  
 RESULT 3  
 A27307  
 proline-rich phosphoprotein (gene PRH1, Db allele) - human  
 A/Alternate names: salivary acidic proline-rich protein  
 C/Species: Homo sapiens (man)

C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 29-Aug-1997  
 C;Accession: A27307  
 R;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
 A;Reference number: A27307; MUID:88074309; PMID:3687941  
 A;Accession: A27307  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 1-171 <AZE>  
 A;Cross-references: EMBL:K03203  
 C;Genetics:  
 A;Gene: GDB:PRH1  
 A;Cross-references: GDB:119515; OMIM:169730  
 A;Map position: 12p13.2-12p13.2  
 C;Superfamily: proline-rich protein  
 C;Keywords: phosphoprotein

Query Match 100.0%; Score 62; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 0.021;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 10  
 |||||  
 DB 120 GGHPRPPRG 129

RESULT 4  
 A24026  
 erythromycin resistance protein - *Arthrobacter* sp.  
 C;Species: *Arthrobacter* sp.  
 C;Date: 22-Jul-1987 #sequence\_revision 22-Jul-1987 #text\_change 18-Jun-1999  
 C;Accession: A24026  
 R;Roberts, A.N.; Hudson, G.S.; Brenner, S.  
 Gene 35, 259-270, 1995  
 A;Reference number: A24026; MUID:86006275; PMID:4043733  
 A;Accession: A24026  
 A;Molecule type: DNA  
 A;Residues: 1-340 <ROB>  
 A;Cross-references: GB:ML1276; NID:G142203; PIDN:AAA22075.1; PID:G142204  
 C;Superfamily: rRNA (adenine-N6-) methyltransferase  
 C;Keywords: antibiotic resistance

Query Match 79.0%; Score 49; DB 2; Length 340;  
 Best Local Similarity 88.9%; Pred. No. 3.6;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 9  
 |||||  
 DB 329 GGRPRPPRG 337

RESULT 5  
 B72654  
 Probable histidyl-tRNA synthetase APE0662 - *Aeropyrum pernix* (strain K1)  
 C;Species: *Aeropyrum pernix*  
 C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 28-Jul-2000  
 C;Accession: B72654  
 R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
 DNA Res. 6, 83-101, 1999  
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyr*  
 A;Reference number: A72450; MUID:99310339; PMID:10382966  
 A;Accession: B72654  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-438 <KAW>  
 A;Cross-references: DDBJ:AP000060; NID:G5104188; PIDN:BAA79634.1; PID:G5104319  
 A;Experimental source: strain K1  
 C;Genetics:  
 A;Gene: APE0662  
 C;Superfamily: histidine-tRNA ligase; histidine-tRNA ligase homology

Query Match 79.0%; Score 49; DB 2; Length 438;  
 Best Local Similarity 88.9%; Pred. No. 4.6;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 9  
 |||||  
 DB 6 GGRPRPPRG 14

RESULT 6  
 AC0398  
 pseudouridylylate synthase (EC 4.2.1.70) [imported] - *Yersinia pestis* (strain CO92)  
 C;Species: *Yersinia pestis*  
 C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
 C;Accession: AC0398  
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.  
 deno-farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.  
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell  
 Nature 413, 523-527, 2001  
 A;Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
 A;Reference number: AB0001; MUID:21470413; PMID:11586360  
 A;Accession: AC0398  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-325 <KUR>  
 A;Cross-references: GB:AL590842; PIDN:CAC92511.1; PID:G15981211; GSPDB:GN00175  
 C;Genetics:  
 A;Gene: rliuD  
 C;Superfamily: conserved hypothetical protein HI0176  
 C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 74.2%; Score 46; DB 2; Length 325;  
 Best Local Similarity 77.8%; Pred. No. 9.7;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 9  
 |||||  
 DB 254 GGRPRPPRG 262

RESULT 7  
 E65037  
 ftsH suppressor protein SfhB - *Escherichia coli* (strain K-12)  
 C;Species: *Escherichia coli*  
 C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
 C;Accession: E65037  
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A;Title: The complete genome sequence of *Escherichia coli* K-12.  
 A;Reference number: A64720; MUID:197426617; PMID:9278503  
 A;Accession: E65037  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-326 <BLAT>  
 A;Cross-references: GB:AE000346; GB:U00096; NID:G2367141; PIDN:AAC75643.1; PID:G178894  
 A;Experimental source: strain K-12, substrain M31655  
 C;Genetics:  
 A;Gene: sfhB  
 C;Superfamily: conserved hypothetical protein HI0176

Query Match 74.2%; Score 46; DB 2; Length 326;  
 Best Local Similarity 77.8%; Pred. No. 9.7;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 9  
 |||||  
 DB 254 GGRPRPPRG 262

RESULT 8  
 A91061  
 suppressor of ftsH mutation [imported] - *Escherichia coli* (strain O157:H7, substrain R



C;Accession: E70610  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: E70610  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-561 <COL>  
A;Cross-references: GB:Z93777; GB:AL123456; NID:G3261726; PIDN:CAB07817.1; PID:e311160  
A;Experimental source: strain H37Rv  
C;Genetics:  
A;Gene: Rv1215c

Query Match 71.0%; Score 44; DB 2; Length 561;  
Best Local Similarity 87.5%; Pred. No. 32;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 HRPDRGR 10  
:|||||  
DB 534 YPRPDRGR 541

## RESULT 13

A37860  
calcium channel protein alpha-1 chain, skeletal muscle - common carp  
C/Species: Cyprinus carpio (common carp)  
C/Date: 31-May-1991 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C/Accession: A37860  
R/Grabner, M.; Friedrich, K.; Knaus, H.G.; Striessnig, J.; Scheffauer, F.; Staudinger, F.  
Proc. Natl. Acad. Sci. U.S.A. 88, 727-731, 1991  
A/Title: Calcium channels from Cyprinus carpio skeletal muscle.  
A/Reference number: A37860; MUID:91126068; PMID:1846962  
A/Accession: A37860  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA

A/Residues: 1-1852 <GRA>  
A/Cross-references: GB:M62554; GB:M37203; NID:G213049; PIDN:AAA49205.1; PID:G213050  
C/Superfamily: voltage-dependent calcium channel protein alpha-1 chain  
C/Keywords: glycoprotein; phosphoprotein; skeletal muscle; transmembrane protein  
F/74-90/Domain: transmembrane #status predicted <TR01>  
F/108-131/Domain: transmembrane #status predicted <TR02>  
F/140-158/Domain: transmembrane #status predicted <TR03>  
F/212-234/Domain: transmembrane #status predicted <TR05>  
F/328-350/Domain: transmembrane #status predicted <TR06>  
F/448-466/Domain: transmembrane #status predicted <TR07>  
F/484-501/Domain: transmembrane #status predicted <TR08>  
F/514-530/Domain: transmembrane #status predicted <TR09>  
F/577-596/Domain: transmembrane #status predicted <TR11>  
F/650-676/Domain: transmembrane #status predicted <TR12>  
F/817-834/Domain: transmembrane #status predicted <TR13>  
F/853-870/Domain: transmembrane #status predicted <TR14>  
F/884-901/Domain: transmembrane #status predicted <TR15>  
F/947-966/Domain: transmembrane #status predicted <TR17>  
F/1057-1084/Domain: transmembrane #status predicted <TR18>  
F/1135-1153/Domain: transmembrane #status predicted <TR19>  
F/1169-1188/Domain: transmembrane #status predicted <TR20>  
F/1197-1215/Domain: transmembrane #status predicted <TR21>  
F/1231-1310/Domain: transmembrane #status predicted <TR23>  
F/1377-1402/Domain: transmembrane #status predicted <TR24>  
F/99,102,274,470,813,1157,1269,1485,1703,1713,1745,1760,1848/binding site: carbohydrate  
F/407/Binding site: phosphate (Thr) (covalent) (by CAMP-dependent kinase) #status predicted  
F/1471,1523,1738/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status

Query Match 71.0%; Score 44; DB 2; Length 1852;

Best Local Similarity 87.5%; Pred. No. 99; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRPPR 8

Db 41 GGNPRPPR 48

## RESULT 14

A46302  
PTB-associated splicing factor, long form - human  
N/Alternate names: myoblast cell surface antigen 24.1D5; polypyrimidine tract-binding protein  
C/Species: Homo sapiens (man)  
C/Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 24-Sep-1999  
C/Accession: A46302; A43557; S29995  
R/Patton, J.G.; Porro, E.B.; Galceran, J.; Tempst, P.; Nadal-Ginard, B.  
Genes Dev. 7, 393-406, 1993  
A/Title: Cloning and characterization of PSF, a novel pre-mRNA splicing factor.  
A/Reference number: A46302; MUID:93194059; PMID:8449401  
A/Accession: A46302

A/Molecule type: mRNA

A/Residues: 1-707 <PAT>

A/Cross-references: EMBL:X70944; NID:G38457; PIDN:CAA50283.1; PID:G38458

A/Note: sequence extracted from NCBI backbone (NCBIP:127206)

R/Gower, H.J.; Moore, S.E.; Dickson, G.; Elsom, V.L.; Nayak, R.; Walsh, F.S.

Development 105, 723-731, 1989

A/Title: Cloning and characterization of a myoblast cell surface antigen defined by 24.1

A/Reference number: A43557; MUID:90091812; PMID:2480877

A/Accession: A43557

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 312-707 <COM>  
A/Cross-references: GB:X16850; NID:G23711; PIDN:CAA34747.1; PID:G23712  
C/Genetics:

A/Gene: GDB:SFPO; PSF

A/Cross-references: GDB:138275

A/Map position: 4q-4q

C/Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein  
C/Keywords: alternative splicing; pre-mRNA splicing; surface antigen  
F/298-359/Domain: ribonucleoprotein repeat homology <RRM1>  
F/372-438/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 70.2%; Score 43.5; DB 2; Length 707;

Best Local Similarity 80.0%; Pred. No. 48;

Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 1 GGHPRPPR-RG 9

Db 228 GGHPRPPHRG 237

## RESULT 15

VGEBHB

glycoprotein gIII precursor - bovine herpesvirus 1 (strain Cooper)

C/Species: bovine herpesvirus 1

C/Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 16-Jul-1999

C/Accession: A32593

R/Ritzpatrick, D.R.; Babiuk, L.A.; Zamb, T.J.

Virology 173, 46-57, 1989

A/Title: Nucleotide sequence of bovine herpesvirus type 1 glycoprotein gIII, a structural  
protein of other herpesviruses.

A/Reference number: A32593; MUID:90051093; PMID:2554578

A/Accession: A32593

A/Molecule type: DNA

A/Residues: 1-521 <FIT>

A/Cross-references: GB:M27491; NID:G330754; PIDN:AAA46054.1; PID:G330755

C/Superfamily: herpesvirus glycoprotein F

F/1-21/Domain: signal sequence #status predicted <SIG>

F/22-521/Product: glycoprotein gIII #status predicted <GPG>

F/467-500/Domain: transmembrane #status predicted <TMN>

F/93,111,164,208/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 69.4%; Score 43; DB 1; Length 521;

Best Local Similarity 87.5%; Pred. No. 42;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGHPRPPR 8

Db 121 GGRPRPPR 128

Search completed: April 6, 2004, 16:16:52

Job time: 12.215 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 6.4486 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-7

Perfect score: 62

Sequence: 1 GGHPRPRGR 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	166	1 PRPC_HUMAN	P02810 homo sapien
2	49	79.0	340	1 ERMA_ARTS3	P09891 arthrobacte
3	49	79.0	438	1 SYH_AERPE	Q9yeb2 aeropyrum p
4	46	74.2	325	1 RLUD_ECO57	Q8x9f0 escherichia
5	46	74.2	325	1 RLUD_ECOL6	Q8fz90 escherichia
6	46	74.2	325	1 RLUD_ECOL1	P33643 escherichia
7	46	74.2	325	1 RLUD_HAEDU	Q91747 haemophilus
8	46	74.2	325	1 RLUD_SALTY	Q8xg92 salmonella
9	46	74.2	325	1 RLUD_YERPE	Q82bv7 yersinia pe
10	44	71.0	1852	1 CCAS_CYPCA	P22316 cyprinus ca
11	43.5	70.2	707	1 SFPO_HUMAN	P23246 homo sapien
12	43	69.4	521	1 VGLC_HSVBC	P14378 bovine herp
13	43	69.4	1596	1 MAM_DROME	P21519 drosophila
14	42	67.7	453	1 A2AB_RAT	P19328 rattus norv
15	42	67.7	455	1 A2AB_MOUSE	P30545 mus musculu
16	41	66.1	414	1 YS2B_CARBL	Q09166 caenorhabdi
17	41	66.1	478	1 Z207_HUMAN	Q43670 homo sapien
18	41	66.1	644	1 ARNT_DROME	O15945 drosophila
19	41	66.1	687	1 WRK2_ARATH	Q19g77 arabidopsis
20	41	66.1	924	1 IF2_MYCLE	P29519 mycobacteri
21	40	64.5	107	1 CVT2_ORYSA	P20907 oryza sativ
22	40	64.5	146	1 VAT7_SCHPO	O10152 schizosacch
23	40	64.5	324	1 RLUD_HAETH	P44445 haemophilus
24	40	64.5	324	1 RLUD_PASMU	Q98ka6 baeturella
25	40	64.5	378	1 LFNG_MOUSE	O09010 mus musculu
26	40	64.5	378	1 LFNG_MOUSE	Q924t4 rattus norv
27	40	64.5	418	1 Y104_AERPE	Q9yazi aeropyrum p
28	40	64.5	440	1 CU63_MOUSE	P59659 mus musculu
29	40	64.5	1172	1 TRP2_HUMAN	Q92444 mus musculu
30	39	62.9	343	1 ALX3_MOUSE	O95076 homo sapien
31	39	62.9	343	1 ALX3_MOUSE	Q70137 mus musculu
32	39	62.9	576	1 Z384_HUMAN	O8cf68 homo sapien
33	39	62.9	579	1 Z384_RAT	Q9eqj4 rattus norv

## RESULT 1

ID	PRPC_HUMAN	STANDARD;	PRT;	166 AA.
AC	P02810;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains: Peptide P-C].			
DE	PRH1 AND PRH2.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC				
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).			
RX	MEDLINE=86196106; PubMed=3009472;			
RA	Kim H.-S., Maeda N.;			
RT	"Structures of two HaeIII-type genes in the human salivary proline-rich protein multigene family.;"			
RL	J. Biol. Chem. 261:6712-6718(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).			
RX	MEDLINE=85289325; PubMed=2993301;			
RA	Maeda N., Kim H.-S., Azen E.A., Smithies O.;			
RT	"Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.;"			
RL	J. Biol. Chem. 260:11123-11130(1985).			
RN	[3]			
RP	SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-P AND PIF-S).			
RX	MEDLINE=89061650; PubMed=3196309;			
RA	Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K.,			
RT	"The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).;"			
RL	Biochem. J. 255:15-21(1988).			
RN	[4]			
RP	SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).			
RX	MEDLINE=88074309; PubMed=3687941;			
RA	Azen E.A., Kim H.S., Goodman P., Flynn S., Maeda N.;			
RT	"Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich proteins Pa, Db, and PIF.;"			
RL	Am. J. Hum. Genet. 41:1035-1047(1987).			
RN	[5]			
RP	SEQUENCE OF 17-166 (PRP-2).			
RX	MEDLINE=88222916; PubMed=3710693;			
RA	Schlesinger D.H., Hay D.I.;			
RT	"Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva.;"			
RL	Int. J. Pept. Protein Res. 27:373-379(1986).			
RN	[6]			
RP	SEQUENCE OF 17-166 (PROTEIN C).			
RX	MEDLINE=80204368; PubMed=7380845;			
RA	Wong R.S.C., Bennick A.;			

34 39 62.9 601 1 3BP1\_MOUSE P55194 mus musculu  
35 39 62.9 618 1 NM24\_MOUSE Q97082 mus musculu  
36 39 62.9 633 1 ROR\_HUMAN O43390 homo sapien  
37 39 62.9 3649 1 ACVS\_NOCOLA P27743 nocardia la  
38 38 61.3 196 1 R15E\_METAC Q8tpx0 methanosarc  
39 38 61.3 196 1 R15E\_METMA Q8btus methanosarc  
40 38 61.3 260 1 DPMI\_MOUSE O70152 mus musculu  
41 38 61.3 266 1 ELMI\_CRIGR Q9wu83 cricetulus  
42 38 61.3 429 1 ELKI\_MOUSE P16969 mus musculu  
43 38 61.3 431 1 UL61\_HCMVA P16818 human cytom  
44 38 61.3 639 1 ZIC5\_HUMAN Q96t25 homo sapien  
45 38 61.3 656 1 P137\_MOUSE Q60865 mus musculu

## ALIGNMENTS

RT "The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A." J. Biol. Chem. 255:5943-5948(1980). [7]

RN SEQUENCE OF 17-46 (PROTEIN C).

RP MEDLINE=81191179; PubMed=7228490;

RA Schlesinger D.H., Hay D.I.;

RT "Primary structure of the active tryptic fragments of human and monkey salivary anionic proline-rich proteins." Int. J. Pept. Protein Res. 17:34-41(1981). [8]

RN SEQUENCE OF 17-122 (PROTEIN A).

RP MEDLINE=79173237; PubMed=438215;

RA Wong R.S.C., Hofmann T., Bennick A.;

RT "The complete primary structure of a proline-rich phosphoprotein from human saliva." J. Biol. Chem. 254:4800-4808(1979). [9]

RN SEQUENCE OF 17-122 (PROTEIN A).

RP Schlesinger D.H., Hay D.I.;

RT "Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent inhibitor of calcium phosphate precipitation in human parotid saliva." (In) Gross E., Meienhofer J. (eds.);

RL Peptides: structure and biological function (Proceedings of the 6th American peptide symposium), pp.133-136, Pierce Chemical Co., Rockford Il. (1979). [10]

RN SEQUENCE OF 123-166 (PEPTIDE P-C).

RP MEDLINE=80227634; PubMed=7390979;

RA Isemura S., Saitoh E., Sanada K.;

RT "The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to a salivary proline-rich phosphoprotein, protein C." J. Biochem. 87:1071-1077(1980). [11]

RN VARIANT PRH2-3 LYS-163.

RP Azen E.A.;

RT "A frequent mutation in the acidic proline-rich protein gene, PRH2, causing a Q147X change closely adjacent to the bacterial binding domain of the cognate salivary PRP (Pr1) in Afro-Americans." Hum. Mutat. 12:72-72(1998).

CC -!- FUNCTION: PRP's act as highly potent inhibitors of crystal growth of calcium phosphates. They provide a protective and reparative environment for dental enamel which is important for the integrity of the teeth.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PRP-S yield PRP-4, PRP-3 (protein A), and PRP-P, respectively.

CC -!- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele; allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is also known as PR1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

FT SIGNAL 1 16  
FT CHAIN 17 166  
FT CHAIN 17 122  
FT CHAIN 123 166  
FT DOMAIN 17 46  
FT MOD\_RES 17 17  
FT MOD\_RES 24 24  
FT MOD\_RES 38 38  
FT VARIANT 20 20  
FT VARIANT 66 66  
FT VARIANT 163 163  
FT CONFLICT 41 41  
SQ SEQUENCE 166 AA; 17017 MW; A7DF62BF94E3C3EF CRC64;

Query Match 100.0%; Score 62; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 0.0071; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0;

QY 1 GGHPRPPRGR 10  
DB 115 GGHPRPPRGR 124

RESULT 2  
ERMA\_ARTS3 STANDARD; PRT; 340 AA.  
ID ERMA\_ARTS3  
AC P09891;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE rRNA adenine N-6-methyltransferase (EC 2.1.1.48) (Macrolide-lincosamide-streptogramin B resistance protein) (Erythromycin resistance protein).  
DE ERMA.  
GN Arthrobacter sp. (Strain B3381).  
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Propionibacteriaceae; Nocardioidaceae; Aeromicrobium.  
OX NCBI\_TaxID=31956;  
RN [1]\_SEQUENCE FROM N.A.  
RX MEDLINE=86006275; PubMed=4043733;  
RA Roberts A.N., Hudson G.S., Brenner S.;

RT "An erythromycin-resistance gene from an erythromycin-producing strain of *Arthrobacter* sp." Gene 35:259-270(1985).

RL -!- FUNCTION: THIS PROTEIN PRODUCES A DIMETHYLATION OF THE ADENINE RESIDUE AT POSITION 2058 IN 23S RNA, RESULTING IN REDUCED AFFINITY BETWEEN RIBOSOMES & MACROLIDE-LINCOSAMIDE-STREPTOGRAMIN B ANTIOTICS.

CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-homocysteine + rRNA containing N(6)-methyladenine.

CC -!- SIMILARITY: Belongs to the rRNA adenine N-6-methyltransferase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M11276; AAA22075.1; --  
DR InterPro: IPR001737; RNA\_A\_dimeth.  
DR InterPro: IPR000051; SAM\_Bind.  
DR Pfam; PF00398; Rnaad; 1.  
DR SMART; SMO0650; RADC; 1.

DR PROSITE; PS01131; RNA\_A\_DIMETH; 1.  
KW Antibiotic resistance; Transferase; Methyltransferase.  
SQ SEQUENCE 340 AA; 37453 MW; E99A714C391952B5 CRC64;

Query Match 79.0%; Score 49; DB 1; Length 340;  
Best Local Similarity 88.9%; Pred. No. 1.4; Gaps 0;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGRPRPRG 9  
|||  
DB 329 GGRPRPRG 337

RESULT 3  
ID SYH\_AERPE STANDARD; PRT; 438 AA.  
AC QSYEB2;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE Histidyl-tRNA synthetase (EC 6.1.1.21) (Histidine--tRNA ligase)  
DE (HisRS).  
GN HISS OR APE0662.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
OC Desulfurococcales; Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KJ;  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,  
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).

CC -1- CATALYTIC ACTIVITY: ATP + L-histidine + tRNA(His) = AMP +  
CC diphosphate + L-histidyl-tRNA(His).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP000060; BAA79634.1; -.  
CC DR PIR; B72654; B72654.  
CC DR HSSP; O32422; 1QE0.  
CC DR HAMAP; MF 00127; -; 1.  
CC DR InterPro; IPR004154; HGTP\_anticonodon.  
CC DR InterPro; IPR004516; HisS.  
CC DR InterPro; IPR002314; tRNA-synt\_2b.  
CC DR InterPro; IPR006195; tRNA\_anticonodon; 1.  
CC DR Pfam; PF03129; HGTP\_anticonodon; 1.  
CC DR Pfam; PF00587; tRNA-synt\_2b; 1.  
CC DR TIGRFAMs; TIGR00442; HisS; 1.  
CC DR PROSITE; PS00862; AA tRNA LIGASE II; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
SQ SEQUENCE 438 AA; 49111 MW; ECCE67F59A9FB7BC CRC64;

Query Match 79.0%; Score 49; DB 1; Length 438;  
Best Local Similarity 88.9%; Pred. No. 1.9;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGRPRPRG 9  
|||  
DB 6 GGRPRPRG 14

RESULT 4  
ID RLUD\_ECO57 STANDARD; PRT; 325 AA.  
AC QX39F0;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)  
DE (Pseudouridylate synthase) (Uracil hydrolyase).  
GN RLUD OR SFHB OR Z3898 OR ECS3457.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074335; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533 (2001).

CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC 5'-phosphate + H(2)O.  
CC -1- SIMILARITY: Belongs to the pseudouridine synthase rluA family.  
CC -1- SIMILARITY: Contains 1 S4 RNA-binding domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP005490; AAG57707.1; -.  
CC DR EMBL; AP002562; BAB36880.1; -.  
CC DR PIR; A91061; A91061.  
CC DR InterPro; IPR006225; Pseud\_synth\_RluD.  
CC DR InterPro; IPR006145; PseudoU\_synth.  
CC DR InterPro; IPR006224; Rlu\_synth.  
CC DR InterPro; IPR002942; S4.  
CC DR Pfam; PF00849; Pseudou\_synth\_2; 1.  
CC DR Pfam; PF01479; S4; 1.  
CC DR ProDom; PD001819; PSI\_RLU; 1.  
CC DR SMART; SM00363; S4; 1.  
CC DR TIGRFAMs; TIGR00005; rluD\_subfam; 1.  
CC DR PROSITE; PS11129; PSI\_RLU; 1.  
CC DR PROSITE; PS50889; S4; 1.  
KW RNA processing; Lyase; RNA-binding; Complete proteome.

```

FT INIT MET 0 0 BY SIMILARITY.
FT DOMAIN 17 90 S4 RNA-BINDING.
FT ACT SITE 138 138 BY SIMILARITY.
SQ SEQUENCE 325 AA; 36990 MW; 7B046623A458967 CRC64;

Query Match 74.2%; Score 46; DB 1; Length 325;
Best Local Similarity 77.8%; Pred. No. 3.9;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 1 GGHPRPPRG 9
Db 253 GGRPRPPKG 261

RESULT 5
RLUD_ECOLI6 STANDARD; PRT; 325 AA.
ID -RLUD_ECOLI6 STANDARD; PRT; 325 AA.
AC Q8FEZ3; O83010; (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
DE (Pseudouridylylate synthase) (Uracil hydrolyase).
GN RLUD OR SFHB OR C3116 OR SF2653 OR S2829.
OS Escherichia coli O6, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=2238234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz F.R.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786 (2003).
CC -!- FUNCTION: Responsible for synthesis of pseudouridine from uracil
at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By
similarity).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
5'-phosphate + H(2)O.
CC -!- SIMILARITY: Belongs to the pseudouridine synthase rluA family.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016764; AAN81585.1; -
CC EMBL; AE015281; AAN44149.1; -
CC EMBL; AE016987; AAP17972.1; -
CC InterPro; IPR006225; Pseud_synth_RLUD.
CC InterPro; IPR006145; Pseudou_synth.
CC InterPro; IPR006224; Rlu_synth.
CC InterPro; IPR002942; S4.
CC Pfam; PF00849; Pseudou_synth_2; 1.
CC Pfam; PF01479; S4; 1.
CC ProDom; PD001819; PSI_RLU; 1.
CC SMART; SM00363; S4; 1.
CC TIGRPFAMs; TIGR00005; rluD_subfam; 1.
CC PROSITE; PS01129; PSI_RLU; 1.
CC PROSITE; PS00889; S4; 1.
CC TrRNA processing; Lyase; RNA-binding; Complete proteome.
FT INIT MET 0 0 BY SIMILARITY.
FT DOMAIN 17 90 S4 RNA-BINDING.
FT ACT SITE 138 138 BY SIMILARITY.
SQ SEQUENCE 325 AA; 36976 MW; 35f58627C17E696C CRC64;

Query Match 74.2%; Score 46; DB 1; Length 325;
Best Local Similarity 77.8%; Pred. No. 3.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 9
Db 253 GGRPRPPKG 261

RESULT 6
RLUD_ECOLI STANDARD; PRT; 325 AA.
ID -RLUD_ECOLI STANDARD; PRT; 325 AA.
AC P33643; P77003;
DT 01-FEB-1994 (Rel. 28, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
DE (Pseudouridylylate synthase) (Uracil hydrolyase).
GN RLUD OR SFHB OR B5594.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Ogura T., Tomoyasu T.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGI655;
RX MEDLINE=97426637; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saio N., Saioh Y., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,

```

RA Yanagata S., Horiuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the *Escherichia coli*  
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features.";  
RN DNA Res. 4:91-113(1997).  
[4]  
RP SEQUENCE OF 39-325 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=91294165; PubMed=1906060;  
RA Kitagawa M., Wada C., Yoshioka S., Yura T.;  
RT "Expression of ClpB, an analog of the ATP-dependent protease -  
RT regulatory subunit in *Escherichia coli*, is controlled by a heat shock  
RT sigma factor (sigma 32)." ;  
RN J. Bacteriol. 173:4247-4253(1991).  
[5]  
RN CHARACTERIZATION.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=99029898; PubMed=9814761;  
RA Raychaudhuri S., Conrad J., Hall B.G., Ofengand J.;  
RT "A pseudouridine synthase required for the formation of two  
RT universally conserved pseudouridines in ribosomal RNA is essential for  
RT normal growth of *Escherichia coli*." ;  
RN RNA 4:1407-1417(1998).  
[6]  
RN FUNCTION, AND SEQUENCE OF N-TERMINUS.  
RC STRAIN=20537953; PubMed=11087118;  
RX Wziesinski J., Bakin A., Ofengand J., Lane B.G.;  
RA "Isolation and properties of *Escherichia coli* 23S-RNA pseudouridine  
RT 1911, 1915, 1917 synthase (RluD)." ;  
RN IUBMB Life 50:33-37(2000).  
[7]  
RN MUTAGENESIS OF ASP-138.  
RC MEDLINE=21345619; PubMed=11453071;  
RX Gutgesell N.S., Del Campo M., Raychaudhuri S., Ofengand J.;  
RA "A second function for pseudouridine synthases: A point mutant of RluD  
RT unable to form pseudouridines 1911, 1915, and 1917 in *Escherichia coli*  
RT 23S ribosomal RNA restores normal growth to an RluD-minus strain." ;  
RN RNA 7:990-998(2001).  
RC -1- FUNCTION: Responsible for synthesis of pseudouridine from uracil  
CC at positions 1911, 1915 and 1917 in 23S ribosomal RNA.  
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC 5'-phosphate + H(2)O.  
CC -1- SIMILARITY: Belongs to the pseudouridine synthase rluA family.  
CC -1- SIMILARITY: Contains 1 S4 RNA-binding domain.  
CC -1- CAUTION: Ref.4 sequence differs from that shown due to a  
CC frameshift in position 133.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U50134; AAA92957.1; -.  
CC EMBL; A9000346; AAC75643.1; -.  
CC EMBL; D90887; BAA16479.1; ALT INIT.  
CC EMBL; X57620; -; NOT\_ANNOTATED\_CDS.  
CC PIR; E65037; E65037.  
CC EcoGene; EG12098; rluD  
CC InterPro; IPR008225; Pseud\_synth\_RluD.  
CC InterPro; IPR006145; Pseudou\_synth.  
CC InterPro; IPR006224; Rlu\_synth.  
CC InterPro; IPR002942; S4.  
CC Pfam; PF00849; Pseudou\_synth\_2; 1.  
CC Pfam; PF01479; S4; 1.  
CC ProDom; PD001819; PSI\_RLU; 1.  
CC SMART; SM00363; S4; 1.  
CC TIGRFAMs; TIGR00005; rluD\_subfam; 1.  
CC PROSITE; PS01129; PSI\_RLU; 1.  
CC PROSITE; PS00889; S4; 1.  
CC rRNA processing; Lyase; RNA-binding; Complete proteome.

FT INIT MET 0 0  
FT DOMAIN 17 90  
FT ACT SITE 138 138  
FT MUTAGEN 138 138  
FT CONFLICT 270 325  
D->N,T: LOSS OF ACTIVITY.  
LRKFDQALHATMLRLYHPISGIEHMEWHAPIQDMVELIEV  
MRADFEHKKDEVDWL -> AGVSLTARRYMQPCCVFTRSP  
ASKWNGMLPHKIWS (IN REF. 3).  
SQ SEQUENCE 325 AA; 36990 MW; 02BDF2AEA4E8300D CRC64;  
Query Match 74.2%; Score 46; DB 1; Length 325;  
Best Local Similarity 77.8%; Pred. No. 3.9;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGHPRPPRG 9  
DB 253 GGRPPPKG 261  
RESULT 7  
RLUD HAEDU STANDARD; PRT; 325 AA.  
AC Q9L7A7;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)  
DE (pseudouridylylate synthase) (Uracil hydrolyase).  
GN RLUD OR HD0469.  
OS Haemophilus ducreyi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=730;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=35000HP / ATCC 700724;  
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,  
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;  
RT "The complete genome sequence of *Haemophilus ducreyi*." ;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE OF 1-250 FROM N.A.  
RC STRAIN=35000HP / ATCC 700724;  
RX MEDLINE=20300369; PubMed=10735974;  
RA Sun S., Schilling B., Tarantino L., Tullius M.V., Gibson B.W.,  
RA Munson R.S. Jr.;  
RT "Cloning and characterization of the lipooligosaccharide  
RT galactosyltransferase II gene of *Haemophilus ducreyi*." ;  
RL J. Bacteriol. 182:2292-2298(2000).  
CC -1- FUNCTION: Responsible for synthesis of pseudouridine from uracil  
CC at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC 5'-phosphate + H(2)O.  
CC -1- SIMILARITY: Belongs to the pseudouridine synthase rluA family.  
CC -1- SIMILARITY: Contains 1 S4 RNA-binding domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; AE017152; AAP95428.1; -.  
CC EMBL; AF224466; AAF32394.1; -.  
CC InterPro; IPR006145; Pseudou\_synth.  
CC InterPro; IPR006224; Rlu\_synth.  
CC InterPro; IPR002942; S4\_synth.  
CC Pfam; PF00849; Pseudou\_synth\_2; 1.  
CC Pfam; PF01479; S4; 1.  
CC ProDom; PD001819; PSI\_RLU; 1.  
CC SMART; SM00363; S4; 1.



```

DR PROSITE; PS01129; PSI_RLU; 1.
DR PROSITE; PS00889; S4; 1.
KW rRNA processing; Lyase; RNA-binding; Complete proteome.
FT DOMAIN 18 78 S4 RNA-BINDING.
FT ACT_SITE 139 139 BY SIMILARITY.
SQ SEQUENCE 325 AA; 36990 MW; A93C172718672DE7 CRC64;

Query Match 74.2%; Score 46; DB 1; Length 325;
Best Local Similarity 77.8%; Pred. No. 3.9;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 1 GGHPRPPRG 9
Db 254 GGRPRPKG 262

RESULT 8
RLUD_SALTY STANDARD; PRT; 325 AA.
ID RLUD_SALTY STANDARD; PRT; 325 AA.
AC Q8XGG2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
DE (pseudouridylylate synthase) (Uracil hydrolyase).
GN RLUD OR SFHB OR STM2662 OR STY2851 OR T2618.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601;

RN SPECIES FROM N.A.
RC SPECIES=S.typhimurium; SPRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porvollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking I., Nhan M.,
Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RL Nature 413:852-856(2001).

RN SPECIES FROM N.A.
RC SPECIES=S.typhi; SPRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Haiden M.T.G., Wain J.,
Churcher C., Mungall K.L., Bentley S.D., White N., Farrar J.,
Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).

RN SPECIES FROM N.A.
RC SPECIES=S.typhi; SPRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18."
RL J. Bacteriol. 185:2330-2337(2003).

CC -!- FUNCTION: Responsible for synthesis of pseudouridine from uracil
at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By
similarity).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
5'-phosphate + H(2)O.
CC -!- SIMILARITY: Belongs to the pseudouridine synthase rluA family.

```

---

```

CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008821; AAL21552.1; -.
CC EMBL; AE627276; CAD05842.1; -.
CC EMBL; AE016842; AA070189.1; -.
DR StyGene; SG????; rluD.
DR InterPro; IPR006225; Pseud synth RluD.
DR InterPro; IPR006145; Pseudou synth.
DR InterPro; IPR006224; Rlu synth.
DR InterPro; IPR003942; S4 synth.
DR Pfam; PF00849; S4; 1.
DR Pfam; PF01479; S4; 1.
DR ProDom; PD001819; PSI_RLU; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00005; rluD subfam; 1.
DR PROSITE; PS01129; PSI_RLU; 1.
DR PROSITE; PS00889; S4; 1.
KW rRNA processing; Lyase; RNA-binding; Complete proteome.
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 17 90 S4 RNA-BINDING.
FT ACT_SITE 138 138 BY SIMILARITY.
SQ SEQUENCE 325 AA; 36969 MW; 7AFDD9A1EDF91D9 CRC64;

Query Match 74.2%; Score 46; DB 1; Length 325;
Best Local Similarity 77.8%; Pred. No. 3.9;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 1 GGHPRPPRG 9
Db 253 GGRPRPKG 261

RESULT 9
RLUD_YERPE STANDARD; PRT; 325 AA.
ID RLUD_YERPE STANDARD; PRT; 325 AA.
AC Q82B7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
DE (pseudouridylylate synthase) (Uracil hydrolyase).
GN RLUD OR SFHB OR YPO3277 OR Y0912.
OS Versinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=632;

RN SPECIES FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
Prentice M.B., Sebahia M., James K.D., Brooks K., Cerdano-Tarraga A.M.,
Baker S., Basham D., Bentley S.D., Davies R.M., Davis P., Dougan G.,
Chillingworth T., Cronin A., Davies R.M., Holroyd S., Jagels K., Karlyshev A.V.,
Feltwell T., Hamlin N., Holroyd S., Oyston P.C.F., Quail M.A., Rutherford K.,
Leather S., Moule S., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Genome sequence of Versinia pestis, the causative agent of plague."
RT Nature 413:523-527(2001).
RL Nature 413:523-527(2001).

RN SPECIES FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,

```



RA Fetherston J.D., Lindler L.E., Brubaker R.R., Piano G.V.,  
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,  
RA Perry R.D.;  
RT "Genome sequence of Yersinia pestis KIM.";  
RL J. Bacteriol. 184:4601-4611 (2002).  
CC -!- FUNCTION: Responsible for synthesis of pseudouridine from uracil  
CC at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By  
CC similarity).  
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC + H<sub>2</sub>O.  
CC -!- SIMILARITY: Belongs to the pseudouridine synthase rluA family.  
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AJ414156; CAC92511.1; -;  
CC EMBL; AE013693; AAM4494.1; -;  
CC PIR; AC0398; AC0398.  
CC InterPro; IPR006225; Pseud\_synth\_RluD.  
CC InterPro; IPR006145; Pseudou synth.  
CC InterPro; IPR006224; Rlu\_synth.  
CC InterPro; IPR002942; S4\_synth.  
CC Pfam; PF00849; Pseudou\_synth\_2; 1.  
CC Pfam; PF01479; S4; 1.  
CC ProDom; PD001819; PSI\_RLUJ; 1.  
CC TIGRFAMs; TIGR00005; RluD\_subfam; 1.  
CC PROSITE; PS01129; PSI\_RLUJ; 1.  
CC PROSITE; PS00889; S4; 1.  
CC rRNA processing; Lyase; RNA-binding; Complete proteome.  
KW DOMAIN 18 91 S4 RNA-BINDING.  
FT ACT\_SITE 139 139 By SIMILARITY.  
SQ SEQUENCE 325 AA; 36706 MW; F05DSB6CDD123ADF CRC64;  
  
Query Match 74.2%; Score 46; DB 1; Length 325;  
Best Local Similarity 77.8%; Pred. No. 3,9;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GGHPRPPRG 9  
||| |||||  
Db 254 GGRPPRPKG 262  
  
RESULT 10  
ID CCAS\_CYPCA STANDARD; PRT; 1852 AA.  
AC P22316;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Dihydropyridine-sensitive L-type, skeletal muscle calcium channel  
DE alpha-1 subunit.  
OS Cyprinus carpio (Common carp).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Cyprinus.  
OX NCBI\_TaxID=7962;  
RN [1]\_TaxID=7962;  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=91126068; PubMed=1846962;  
RA Grabner M., Friedrich K., Knaus H.-G., Striessnig J., Scheffauer F.,  
RA Staudinger R., Koch W.J., Schwartz A., Glossmann H.;  
RT "Calcium channels from Cyprinus carpio skeletal muscle.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:727-731 (1991).  
CC -!- FUNCTION: Voltage-sensitive calcium channels (VSCC) mediate the  
CC entry of calcium ions into excitable cells and are also involved  
CC in a variety of calcium-dependent processes, including muscle

CC contraction, gene expression, cell motility, cell division and  
CC cell death. The isoform alpha-1S gives rise to L-type calcium  
CC currents, long-lasting (L-type) calcium channels, belong to the  
CC "high-voltage activated" (HVA) group. They are blocked by  
CC dihydropyridines (DHP), phenylalkylamines, benzothiazepines, and  
CC B omega-agatoxin-IIIA (omega-aga-IIIA). They are however  
CC insensitive to omega-conotoxin-GVIA (omega-CTX-GVIA) and omega-  
CC agatoxin-IVA (omega-Aga-IVA). Calcium channels containing the  
CC alpha-1S subunit play an important role in excitation-contraction  
CC coupling in skeletal muscle (By similarity).  
CC -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT  
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS  
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-  
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS  
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM  
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA  
CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY. AN  
CC ADDITIONAL GAMMA SUBUNIT IS PRESENT ONLY IN SKELETAL MUSCLE L-TYPE  
CC CHANNEL (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: Skeletal muscle.  
CC -!- PTM: MAY NOT BE PHOSPHORYLATED.  
CC -!- SIMILARITY: Belongs to the calcium channel alpha-1 subunits  
CC family.  
CC  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; M62554; AAA49205.1; -;  
CC PIR; A37860; A37860.  
CC InterPro; IPR001682; Ca\_Na\_pore.  
CC InterPro; IPR002077; Ca\_channel\_alpha.  
CC InterPro; IPR002111; Cat\_channel\_trpL.  
CC InterPro; IPR005821; Ion\_trans.  
CC InterPro; IPR005446; LVDCALPHAL.  
CC InterPro; IPR005450; LVDCALPHALS.  
CC InterPro; IPR005820; M-channel\_nlg.  
CC Pfam; PF00520; Ion\_trans\_4.  
CC PRINTS; PR00167; CACHANNEL.  
CC PRINTS; PR01630; LVDCALPHAL.  
CC PRINTS; PR01634; LVDCALPHALS.  
CC Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
KW Calcium channel; Glycoprotein; Repeat; Multigene family;  
KW Calcium-binding; Phosphorylation.  
FT REPEAT 57 354 I.  
FT REPEAT 433 679 II.  
FT REPEAT 802 1084 III.  
FT REPEAT 1121 1405 IV.  
FT DOMAIN 1 70 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 71 86 S1 OF REPEAT I.  
FT DOMAIN 87 107 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 108 127 S2 OF REPEAT I.  
FT DOMAIN 128 139 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 140 155 S3 OF REPEAT I.  
FT DOMAIN 156 176 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 177 195 S4 OF REPEAT I.  
FT DOMAIN 196 214 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 215 234 S5 OF REPEAT I.  
FT DOMAIN 235 326 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 327 351 S6 OF REPEAT I.  
FT DOMAIN 352 447 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 448 466 S1 OF REPEAT II.  
FT DOMAIN 467 481 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 482 501 S2 OF REPEAT II.  
FT DOMAIN 502 509 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 510 528 S3 OF REPEAT II.  
FT DOMAIN 529 538 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 539 557 S4 OF REPEAT II.

FT	DOMAIN	558	576	CYTOPLASMIC (POTENTIAL).	DT
FT	TRANSMEM	577	596	S5 OF REPEAT II.	DT
FT	DOMAIN	597	651	EXTRACELLULAR (POTENTIAL).	DE
FT	TRANSMEM	652	675	S6 OF REPEAT II.	DE
FT	DOMAIN	676	815	CYTOPLASMIC (POTENTIAL).	DE
FT	TRANSMEM	816	834	S1 OF REPEAT III.	GN
FT	DOMAIN	835	850	EXTRACELLULAR (POTENTIAL).	OS
FT	TRANSMEM	851	870	S2 OF REPEAT III.	OC
FT	DOMAIN	871	882	CYTOPLASMIC (POTENTIAL).	OC
FT	TRANSMEM	883	901	S3 OF REPEAT III.	OX
FT	DOMAIN	902	908	EXTRACELLULAR (POTENTIAL).	RN
FT	TRANSMEM	909	927	S4 OF REPEAT III.	RC
FT	DOMAIN	928	946	CYTOPLASMIC (POTENTIAL).	RX
FT	TRANSMEM	947	966	S5 OF REPEAT III.	RA
FT	DOMAIN	967	1056	EXTRACELLULAR (POTENTIAL).	RT
FT	TRANSMEM	1057	1081	S6 OF REPEAT III.	RT
FT	DOMAIN	1082	1134	CYTOPLASMIC (POTENTIAL).	RL
FT	TRANSMEM	1135	1153	S1 OF REPEAT IV.	RN
FT	DOMAIN	1154	1168	EXTRACELLULAR (POTENTIAL).	RP
FT	TRANSMEM	1169	1188	S2 OF REPEAT IV.	RX
FT	DOMAIN	1189	1196	CYTOPLASMIC (POTENTIAL).	RA
FT	TRANSMEM	1197	1215	S3 OF REPEAT IV.	RT
FT	DOMAIN	1216	1252	EXTRACELLULAR (POTENTIAL).	RT
FT	TRANSMEM	1253	1271	S4 OF REPEAT IV.	RL
FT	DOMAIN	1272	1290	CYTOPLASMIC (POTENTIAL).	RP
FT	TRANSMEM	1291	1310	S5 OF REPEAT IV.	RX
FT	DOMAIN	1311	1377	EXTRACELLULAR (POTENTIAL).	RA
FT	TRANSMEM	1378	1402	S6 OF REPEAT IV.	RT
FT	DOMAIN	1403	1852	CYTOPLASMIC (POTENTIAL).	RL
FT	DOMAIN	4	12	POLY-GLY.	RN
FT	DOMAIN	400	403	POLY-LEU.	RP
FT	DOMAIN	577	583	POLY-LEU.	RX
FT	DOMAIN	1062	1068	POLY-ILE.	RA
FT	DOMAIN	374	391	BINDING TO THE BETA SUBUNIT (BY SIMILARITY).	RT
FT	SITE	309	309	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).	CC
FT	SITE	628	628	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).	CC
FT	SITE	1030	1030	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).	CC
FT	SITE	1344	1344	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).	CC
FT	BINDING	1004	1093	DIHYDROPYRIDINES (BY SIMILARITY).	CC
FT	BINDING	1358	1424	DIHYDROPYRIDINES (BY SIMILARITY).	CC
FT	BINDING	1370	1413	PHENYLALKYLAMINES (BY SIMILARITY).	CC
FT	CA_BIND	1431	1442	BY SIMILARITY.	CC
FT	MOD_RES	407	407	PHOSPHORYLATION (BY PKA) (POTENTIAL).	CC
FT	MOD_RES	1413	1413	PHOSPHORYLATION (BY PKA) (POTENTIAL).	CC
FT	MOD_RES	1471	1471	PHOSPHORYLATION (BY PKA) (POTENTIAL).	CC
FT	MOD_RES	1523	1523	PHOSPHORYLATION (BY PKA) (POTENTIAL).	CC
FT	MOD_RES	1738	1738	PHOSPHORYLATION (BY PKA) (POTENTIAL).	CC
FT	CARBOHYD	99	99	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	102	102	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	274	274	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	470	470	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	1157	1157	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	SEQUENCE	1852	210096	MM; 7DSFAE83501D6CDB CRC64;	CC

Query Match 71.0%; Score 44; DB 1; Length 1852;  
Best Local Similarity 87.5%; Pred. No. 47;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGNRPFR 8  
Db 41 GGNRPFR 48

RESULT 11  
SFPO\_HUMAN  
ID SFPO\_HUMAN STANDARD; PRT; 707 AA.  
AC P23246; F30808;  
DT 01-NOV-1991 (Rel. 20, Created)

01-OCT-1996 (Rel. 34, Last sequence update)  
15-MAR-2004 (Rel. 43, Last annotation update)  
Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-  
binding protein-associated splicing factor) (PFB-associated splicing  
factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).  
SFPO OR PSF.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.  
TISSUE=Fetal brain;  
MEDLINE=93194059; PubMed=8449401;  
Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;  
"Cloning and characterization of PSF, a novel pre-mRNA splicing  
factor.";  
Genes Dev. 7:393-406(1993).  
[2]  
SEQUENCE OF 312-707 FROM N.A.  
TISSUE=Fetal skeletal muscle;  
MEDLINE=90091812; PubMed=2480877;  
Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh P.S.;  
"Cloning and characterization of a myoblast cell surface antigen  
defined by 24.1D5 monoclonal antibody.";  
Development 105:723-731(1989).  
[3]  
SEQUENCE OF 48-68 AND 213-246.  
MEDLINE=93176127; PubMed=8439294;  
Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;  
"Purification and characterization of a DNA-binding heterodimer of 52  
and 100 kDa from HeLa cells.";  
Biochem. J. 290:267-272(1993).  
-!- FUNCTION: Essential pre-mRNA splicing factor required early in  
spliceosome formation. Binds to the mammalian polypyrimidine  
tracts. Forms a complex with the polypyrimidine tract-binding  
protein (PTB). Seems to also bind DNA.  
-!- SUBUNIT: Heterotetramer of two 52 kDa and two 100 kDa subunits.  
-!- SUBCELLULAR LOCATION: Nuclear.  
-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Comment=Additional isoforms seem to exist;  
Name=Long;  
IsoId=D33246-1; Sequence=Displayed;  
Name=Short;  
IsoId=D33246-2; Sequence=VSP\_003855;  
-!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.  
-!- CAUTION: Was originally (Ref.2) thought to be myoblast cell  
surface antigen 24.1D5 and a possible membrane-bound protein  
ectokinase.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X70944; CAA50283.1; -.  
EMBL; X16850; CAA34747.1; -.  
PIR; A46302; A46302.  
PIR; S29770; S29770.  
HSSP; P11940; 1CVJ.  
SWISS-2DPAGE; P23246; HUMAN.  
Genew; HGNC:10774; SFPO.  
DR GK; P23246; -.  
DR MIM; 605199; -.  
DR GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.  
DR GO; GO:0006397; P:mRNA processing; TAS.  
DR GO; GO:0006371; P:mRNA splicing; TAS.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 2.

DR	InterPro; IPRO01038; Gp13_EHV.
DR	InterPro; IPR007110; Ig-like.
DR	InterPro; IPR001654; Marek A.
DR	Pfam; PF02124; Marek A; 1.
DR	PRINTS; PR00668; GLYCOPROTEIN.C
DR	PROSITE; PS50835; IG LIKE; 1.
KW	Immunoglobulin domain; Glycoprotein; Transmembrane; Signal.
FT	SIGNAL 1 21
FT	CHAIN 21 521
FT	DOMAIN 32 92
FT	DOMAIN 155 227
FT	DOMAIN 386 451
FT	DOMAIN 467 500
FT	CARBSHM 111 111
FT	CARBOHYD 164 164
FT	CARBOHYD 208 208
FT	SEQUENCE 521 AA; 53584 MW; 72B97266ASABEAS CRC64;
QY	Query Match 69.4%; Score 43; DB 1; Length 521;
DB	Best Local Similarity 87.5%; Pred. No.18;
	Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 GGHRPPR 8
DB	121 GGRPPRP 128
	STANDARD; PRT; 1596 AA.
RESULT 13	
MAM_DROME	
ID	NAM_DROME STANDARD; PRT; 1596 AA.
AC	F21519;
DT	01-MAY-1991 (Rel. 18, Created)
DT	01-MAY-1991 (Rel. 19, Last sequence update)
DT	15-MAR-2004 (Rel. 43, Last annotation update)
DE	Neurogenic protein mastermind.
DE	Neurogenic protein mastermind.
GN	NAM.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
[1]	
RN	
RP	SEQUENCE FROM N.A.
RC	STRAIN=Canton-S;
RX	MEDLINE=91065516; PubMed=1701150;
RA	Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,
RA	Vedvobnick B.;
RT	"The Drosophila neurogenic locus mastermind encodes a nuclear protein
RT	unusually rich in amino acid homopolymers.";
RL	Genes Dev. 4:1688-1700(1990).
CC	-!- FUNCTION: May have a regulatory function possibly in association
CC	with the N gene product.
CC	-!- SUBCELLULAR LOCATION: Nuclear.
CC	-!- DEVELOPMENTAL STAGE: During early neurogenesis mam products are
CC	ubiquitously located. During later stages they accumulate in the
CC	central nervous system.
CC	-!- MISCELLANEOUS: The protein has many AA homopolymeric domains: 21
CC	poly-Gln runs from 5 to 16 AA in length), 4 poly-Gly (6 to 10
CC	AA), 3 poly-Asn (3 x 5 AA), 1 poly-Ala (10 AA) and 1 poly-Thr (5
CC	AA) runs.
CC	-!- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC	YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC	-----
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL; X54251; CAA38152.1 -.
DR	PIR; A36391; A33106

```

DR FlyBase; FBgn002643; man.
DR GO:0005634; C:nucleus; IDA.
DR GO:0007500; P:mesoderm cell fate determination; IMP.
DR GO:0007399; P:neurogenesis; NAS.
KW Neurogenesis; Nuclear protein; Repeat.
FT DOMAIN 20 84
FT DOMAIN 127 190 ARG/LYS-RICH (BASIC).
FT DOMAIN 136 219 GLN-RICH.
FT DOMAIN 229 304 ASN-RICH.
FT DOMAIN 335 388 GLY/ASN-RICH.
FT DOMAIN 392 406 GLN-RICH.
FT DOMAIN 407 440 GLN-RICH.
FT DOMAIN 651 671 GLN-RICH.
FT DOMAIN 700 714 GLN-RICH.
FT DOMAIN 759 816 GLN-RICH.
FT DOMAIN 987 996 S X 2 AA TANDEM REPEATS OF G-V.
FT DOMAIN 1060 1079 ALA-RICH.
FT DOMAIN 1092 1107 8 X 2 AA TANDEM REPEATS OF V-G.
FT DOMAIN 1237 1252 7 X 2 AA TANDEM REPEATS OF G-V.
FT DOMAIN 1492 1496 POLY-THR.
FT DOMAIN 1559 1592 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 1596 AA; 167717 MW; B944D86EF359D605 CRC64;

Query Match 69.4%; Score 43; DB 1; Length 1596;
Best Local Similarity 77.8%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGHPRPRPG 9
Db 639 GGFPRPHG 647

RESULT 14
A2AB RAT
ID A2AB RAT STANDARD; PRT; 453 AA.
AC P19328; Q63021; Q92554;
DT 01-NOV-1990 (Rel. 16, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor).
GN ADRA2B
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90222177; PubMed=2158103;
RA Zeng D., Harrison J.K., D'Angelo D.D., Barber C.M., Tucker A.L.,
RA Lu Z., Lynch K.R.;
RT "Molecular characterization of a rat alpha 2B-adrenergic receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:3102-3106(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Schaak S., Cussac D., Paris H.;
RT "Cloning and characterization of the rat alpha2B-adrenergic receptor
gene promoter.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 6-453 FROM N.A.
RC STRAIN=Sabra; TISSUE=Kidney;
RX MEDLINE=35275432; PubMed=7755946;
RA le Jossec M., Cloix J.F., Pecquery R., Giudicelli Y., Dausse J.P.;
RT "Differential sodium regulation between salt-sensitive and salt-
resistant Sabra rats is not due to any mutation in the renal alpha
2B-adrenoceptor gene.";
RL Am. J. Hypertens. 8:177-182(1995).
CC -1- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-
induced inhibition of adenylyate cyclase through the action of G
proteins.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

```

---

```

CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M32061; AAA40635.1; -.
CC EMBL; AF366899; AA53388.1; -.
CC EMBL; X74400; CAAS2411.1; -.
CC PIR; A35642; A35642.
CC HSSP; P29274; 1MMH.
CC InterPro; IPR000276; GPCR_Rhodpsn.
CC Pfam; PF00001; 7tm.1; 1.
CC PRINTS; PR00237; GPCRHOOPS.
CC PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
CC PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
CC G-protein coupled receptor; Transmembrane; Multigene family;
KW Phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 17 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 18 42 1 (POTENTIAL).
FT DOMAIN 43 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 80 2 (POTENTIAL).
FT DOMAIN 81 90 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 91 113 3 (POTENTIAL).
FT DOMAIN 114 135 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 136 158 4 (POTENTIAL).
FT DOMAIN 159 174 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 175 198 5 (POTENTIAL).
FT DOMAIN 199 375 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 376 399 6 (POTENTIAL).
FT DOMAIN 400 408 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 409 432 7 (POTENTIAL).
FT DOMAIN 433 453 CYTOPLASMIC (POTENTIAL).
FT DISULFID 90 169 BY SIMILARITY.
FT LIPID 445 449 S-palmitoyl cysteine (Potential).
FT DOMAIN 300 314 ASP/GLU-RICH (ACIDIC).
FT SITE 97 97 IMPLICATED IN LIGAND BINDING (BY
SIMILARITY).
FT SITE 181 181 IMPLICATED IN CATECHOL AGONIST BINDING
(BY SIMILARITY).
FT SITE 185 185 IMPLICATED IN CATECHOL AGONIST BINDING
(BY SIMILARITY).
FT CONFLICT 132 132 R -> C (IN REF. 1).
FT CONFLICT 162 163 EP -> DA (IN REF. 1).
SQ SEQUENCE 453 AA; 50369 MW; CBA69CE23EACB511 CRC64;

Query Match 67.7%; Score 42; DB 1; Length 453;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GHPRPPR 8
Db 254 GHPKPPR 260

RESULT 15
A2AB MOUSE
ID A2AB MOUSE STANDARD; PRT; 455 AA.
AC P30545;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor).
GN ADRA2B
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

RP SEQUENCE FROM N.A.  
RC STRAIN=DBA/2; TISSUE=Liver;  
RX MEDLINE=93129625; PubMed=1336396;  
RA Chen W.-M., Chang A.C., Shie B.J., Chang Y.-H., Chang N.-C.A.;  
RT "Molecular cloning and characterization of a mouse alpha 2C2  
RT adrenoceptor subtype gene."  
RL Biochim. Biophys. Acta 1171:219-223(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92378586; PubMed=1354956;  
RA Chruscinski A.J., Link R.E., Daunt D.A., Barsh G.S., Kobilka B.K.;  
RT "Cloning and expression of the mouse homolog of the human alpha 2-C2  
RT adrenergic receptor."  
RL Biochem. Biophys. Res. Commun. 186:1280-1287(1992).  
CC -1- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-  
CC induced inhibition of adenylate cyclase through the action of G  
CC proteins.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M94583; AA473895.1; -  
DR EMBL; L00979; AAA37131.1; ALT\_INIT.  
DR PIR; S28221; S28221.  
DR HSP; P29274; 1MMH.  
DR MGD; MGI:87935; Adra2b.  
DR GO; GO:0004938; F:alpha2-adrenergic receptor activity; IDA.  
DR GO; GO:0001525; P:angiogenesis; IMP.  
DR GO; GO:0000165; P:MAPKK cascade; IMP.  
DR InterPro; IPR000276; GPCR\_Rhodpsn.  
DR Pfam; PF00001; 7tm\_1; 1.  
DR PRINTS; PR00237; GPCRHOPOPSN.  
DR PROSITE; PS00237; G PROTEIN RECP FL 1; 1.  
DR PROSITE; PS0262; G PROTEIN RECP FL 2; 1.  
KW G-protein coupled receptor; Transmembrane; Multigene family;  
KW Phosphorylation; Lipoprotein; Palmitate.  
FT DOMAIN 1 17 EXTRACELLULAR (POTENTIAL).  
FT TRANSNEM 18 42 1 (POTENTIAL).  
FT DOMAIN 43 54 CYTOPLASMIC (POTENTIAL).  
FT TRANSNEM 55 80 2 (POTENTIAL).  
FT DOMAIN 81 90 EXTRACELLULAR (POTENTIAL).  
FT TRANSNEM 91 113 3 (POTENTIAL).  
FT DOMAIN 114 135 CYTOPLASMIC (POTENTIAL).  
FT TRANSNEM 136 158 4 (POTENTIAL).  
FT DOMAIN 159 174 EXTRACELLULAR (POTENTIAL).  
FT TRANSNEM 175 198 5 (POTENTIAL).  
FT DOMAIN 199 377 CYTOPLASMIC (POTENTIAL).  
FT TRANSNEM 378 401 6 (POTENTIAL).  
FT DOMAIN 402 410 EXTRACELLULAR (POTENTIAL).  
FT TRANSNEM 411 434 7 (POTENTIAL).  
FT DOMAIN 435 455 CYTOPLASMIC (POTENTIAL).  
FT DISULFID 90 169 BY SIMILARITY.  
FT LIPID 447 447 S-palmitoyl cysteine (potential).  
FT DOMAIN 306 316 ASP/GLU-RICH (ACIDIC).  
FT SITE 97 IMPLICATED IN LIGAND BINDING (BY  
FT SIMILARITY).  
FT SITE 181 IMPLICATED IN CATECHOL AGONIST BINDING  
FT SITE 185 IMPLICATED IN CATECHOL AGONIST BINDING  
FT CONFLICT 202 202 V -> L (IN REF. 1).  
FT CONFLICT 229 230 MISSING (IN REF. 2).  
SQ SEQUENCE 455 AA; 50615 MW; A3954AD7650E6263 CRC64;

Query Match 67.7%; Score 42; DB 1; Length 455;  
Best Local Similarity 85.7%; Pred. No. 23;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GHPREPR 8  
DB 256 GHPREPR 262  
Search completed: April 6, 2004, 16:07:59  
Job time : 5.4486 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 36.1682 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-7

Perfect score: 62

Sequence: 1 GGHPRPPRGR 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_nhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	77.4	259	16 Q7W1A7	Q7W1A7 bordetella
2	48	77.4	707	4 Q86VG2	Q86VG2 homo sapien
3	46	74.2	621	13 Q7T2J1	Q7T2J1 brachydanio
4	45	72.6	268	10 Q39487	Q39487 coloccasia e
5	45	72.6	407	10 Q8H3U2	Q8H3U2 oryza sativ
6	44	71.0	561	16 Q05316	Q05316 mycobacteri
7	44	71.0	561	16 Q7U0D5	Q7U0D5 mycobacteri
8	44	71.0	646	16 Q98DE3	Q98DE3 rhizobium l
9	43.5	70.2	383	11 Q9ERW2	Q9ERW2 mus musculu
10	43.5	70.2	637	4 Q9BSV4	Q9BSV4 homo sapien
11	43.5	70.2	699	11 Q8VIJ6	Q8VIJ6 mus musculu
12	43	69.4	428	5 Q8SZU8	Q8SZU8 drosophila
13	43	69.4	508	12 Q65821	Q65821 bovine herp
14	43	69.4	537	5 Q86545	Q86545 lymantria d
15	43	69.4	545	2 Q05089	Q05089 nocardioid
16	43	69.4	862	10 Q8W4X6	Q8W4X6 prunus dulc

17	43	69.4	1594	5	Q9V6W7	Q9V6W7 drosophila
18	42	67.7	84	16	Q8DL11	Q8DL11 synchococc
19	42	67.7	174	2	Q8QM5	Q8QM5 saccharopol
20	42	67.7	181	2	Q8KPP8	Q8KPP8 synchococc
21	42	67.7	218	5	Q01331	Q01331 canorhabdi
22	42	67.7	332	2	Q54729	Q54729 synchococc
23	42	67.7	448	11	Q925K7	Q925K7 mus musculu
24	42	67.7	448	11	Q925K6	Q925K6 mus musculu
25	42	67.7	479	16	Q9RRE7	Q9RRE7 deinococcus
26	42	67.7	485	3	Q94115	Q94115 aureobasidi
27	42	67.7	494	5	Q8W777	Q8W777 branchiosto
28	42	67.7	535	10	Q84SD0	Q84SD0 oryza sativ
29	42	67.7	950	11	Q923A9	Q923A9 mus musculu
30	42	67.7	2157	11	Q921R1	Q921R1 mus musculu
31	42	67.7	2158	11	Q7TSC1	Q7TSC1 mus musculu
32	41	66.1	133	10	Q92T22	Q92T22 glycine max
33	41	66.1	322	11	Q7TNE5	Q7TNE5 mus musculu
34	41	66.1	340	4	Q8N395	Q8N395 homo sapien
35	41	66.1	346	4	Q8IU99	Q8IU99 homo sapien
36	41	66.1	381	16	Q8DF67	Q8DF67 vibrio vuln
37	41	66.1	460	11	Q92326	Q92326 mus musculu
38	41	66.1	464	11	Q99LA2	Q99LA2 mus musculu
39	41	66.1	481	16	P73567	P73567 synchocyst
40	41	66.1	495	11	Q9TMD0	Q9TMD0 mus musculu
41	41	66.1	533	10	Q7X188	Q7X188 oryza sativ
42	41	66.1	539	4	Q96IE7	Q96IE7 homo sapien
43	41	66.1	1117	4	Q8NEM6	Q8NEM6 homo sapien
44	41	66.1	1384	2	Q9AJP5	Q9AJP5 myxococcus
45	41	66.1	1406	11	Q80TR1	Q80TR1 mus musculu

## ALIGNMENTS

### RESULT 1

Q7W1A7 PRELIMINARY; PRT; 259 AA.  
ID Q7W1A7  
AC Q7W1A7  
DT 01-OCT-2003 (TREMREL. 25, Created)  
DT 01-OCT-2003 (TREMREL. 25, Last sequence update)  
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)  
DE Hypothetical protein.  
GN BPP0789.  
OS Bordetella parapertussis.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Bordetella.  
OX NCBI\_TaxID=519;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=12822 / ATCC BAA-587;  
RX MEDLINE=22827954; PubMed=12910271;  
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,  
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
RA Leather S., Moule S., Norbertczak H., O'Neil S., Ormond D., Price C.,  
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
RA Unwin L., Whitehead S., Barrall B.G., Maskell D.J.;  
RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
RT Bordetella parapertussis and Bordetella bronchiseptica.";  
RL Nat. Genet. 35:32-40(2003).  
DR EMBL; BX640425; CAB40198.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28165 MW; DFE32C6FBCA19645 CRC64;

Query Match 77.4%; Score 48; DB 16; Length 259;  
Best Local Similarity 80.0%; Pred. No. 8.1;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GGHPRPPRGR 10

```
Db 20 GGEPRPTGR 29
||| ||| ||| ||| |||
RESULT 2
Q86VG2 PRELIMINARY; PRT; 707 AA.
ID Q86VG2; 2003 (TRENBLrel. 24, Created)
AC Q86VG2; 2003 (TRENBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 25, Last annotation update)
DE 12KD storage protein (Fragment).
OS Colocasia esculenta (Elephant's ear) (Taro).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Araceae; Colocasia.
OX NCBI_TaxID=4460;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=IVOT ID No.53; TISSUE=Mature tuber;
RX MEDLINE=9407218; PubMed=8251188;
RA Hirai M., Nakamura K., Imai T., Sato T.;
RT "cDNAs encoding for storage proteins in the tubers of taro (Colocasia
RL esculenta Schott).";
RL Jpn. J. Genet. 68:229-236(1993).
DR EMBL; D16173; BAA03722.1; -.
DR HSSP; P30617; LJPC.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR001480; B_lectin.
DR Pfam; PF01453; Agglutinin; 2.
DR SMART; SM00108; B_lectin; 2.
FT NON_TER 1
FT CHAIN 28 143 12-A PROTEIN.
FT CHAIN 144 268 12-B PROTEIN.
SQ SEQUENCE 268 AA; 29430 MW; DCDE7355788D69FF CRC64;
Query Match 72.4%; Score 48; DB 4; Length 707;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCHPRPPR 8
Db 228 GCHPKPPR 235
||| ||| ||| ||| |||
RESULT 3
Q7T2J1 PRELIMINARY; PRT; 621 AA.
ID Q7T2J1; 2003 (TRENBLrel. 25, Created)
AC Q7T2J1; 2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Membrane-type matrix metalloproteinase 1 beta.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
SEQUENCE FROM N.A.
RC Zhang J., Bai S., Zhang X., Nagase H., Sarvas M.P. Jr.;
RA "The expression of novel membrane-type matrix metalloproteinase
RT isoforms is required for normal development of zebrafish embryos.";
RL Matrix Biol. 0:0-0(2003).
DR EMBL; AY324197; AAP74484.1; -.
SQ SEQUENCE 621 AA; 71129 MW; 6C787905528BDBAB CRC64;
Query Match 74.2%; Score 46; DB 13; Length 621;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCHPRPPR 7
Db 281 GCHPRPPR 287
||| ||| ||| ||| |||
RESULT 4
Q39487 PRELIMINARY; PRT; 268 AA.
ID Q39487; 2003 (TRENBLrel. 01, Created)
AC Q39487; 2003 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TRENBLrel. 25, Last annotation update)
DE 12KD storage protein (Fragment).
OS Colocasia esculenta (Elephant's ear) (Taro).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Araceae; Colocasia.
OX NCBI_TaxID=4460;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=IVOT ID No.53; TISSUE=Mature tuber;
RX MEDLINE=9407218; PubMed=8251188;
RA Hirai M., Nakamura K., Imai T., Sato T.;
RT "cDNAs encoding for storage proteins in the tubers of taro (Colocasia
RL esculenta Schott).";
RL Jpn. J. Genet. 68:229-236(1993).
DR EMBL; D16173; BAA03722.1; -.
DR HSSP; P30617; LJPC.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR001480; B_lectin.
DR Pfam; PF01453; Agglutinin; 2.
DR SMART; SM00108; B_lectin; 2.
FT NON_TER 1
FT CHAIN 28 143 12-A PROTEIN.
FT CHAIN 144 268 12-B PROTEIN.
SQ SEQUENCE 268 AA; 29430 MW; DCDE7355788D69FF CRC64;
Query Match 72.6%; Score 45; DB 10; Length 268;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 GCHPRPPR 8
Db 15 GCHPRPPR 21
||| ||| ||| ||| |||
RESULT 5
Q8H3U2 PRELIMINARY; PRT; 407 AA.
ID Q8H3U2; 2003 (TRENBLrel. 23, Created)
AC Q8H3U2; 2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE P0418E08.7 protein.
GN P0418E08.7.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
RT clone: P0418E08.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004382; BAC16073.1; -.
SQ SEQUENCE 407 AA; 43973 MW; D39E5777F8252C8 CRC64;
Query Match 72.6%; Score 45; DB 10; Length 407;
Best Local Similarity 87.5%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 HPRPRPPR 10
Db 118 HPRPRPPR 125
||| ||| ||| ||| |||
RESULT 6
O05316
```

DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Conserved hypothetical protein.
GN	MBI247C.
OS	Mycobacterium bovis.
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OX	Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1765;	
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=AF2122/97;
RC	MEDLINE=22709107; PubMed=12788972;
RA	Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA	Pryor M., Duchoy S., Grendin S., Lacroix C., Monsempé C., Simon S.,
RA	Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA	Parthill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT	"The complete genome sequence of Mycobacterium bovis";
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RK	EMBL; BX248338; CAD94108.1; -.
KW	Complete proteome.
SQ	SEQUENCE 561 AA; 62626 MW; C35176E8172866AD CRC64;
Query Match	71.0%; Score 44; DB 16; Length 561;
Best Local Similarity	87.5%; Pred.No. 67;
Matches	7; Conservative 1; Mismatches 0; Indels 0; Gaps
QY	3 HRPDPRGR 10
DB	:
	534 YRPDPRGR 541
RESULT 8	
Q98DE3	PRELIMINARY; PRT; 646 AA.
ID	O98DE3
AC	O98DE3;
DT	01-OCT-2001 (TrEMBLrel. 18, Created)
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	2-dehydro-3-deoxygalactonate kinase.
DE	MLR4742.
GN	Rhizobium loti (Mesorhizobium loti).
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC	Phyllobacteriaceae; Mesorhizobium.
OX	NCBI_TaxID=381;
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=MAFF303099;
RC	MEDLINE=21082930; PubMed=11214968;
RA	Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA	Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA	Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA	Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA	Takeuchi C., Yamada M., Tabata S.;
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium
RL	Mesorhizobium loti.";
RL	DNA Res. 7:331-338(2000).
DR	EMBL; AP003004; BAB51328.1; -.
DR	GO; GO:0016301; P:Kinase activity; IEA.
DR	GO; GO:0004601; P:Peroxidase activity; IEA.
DR	GO; GO:0006979; P:Response to oxidative stress; IEA.
DR	InterPro; IPR007729; DGOX.
DR	InterPro; IPR002016; Peroxidase.
DR	InterPro; IPR007087; Znf_C2H2.
DR	Pfam; PF05035; DGOK; 1.
DR	PROSITE; PS00435; PEROXIDASE_1; 1.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW	Kinase; Complete proteome.
SQ	SEQUENCE 646 AA; 68943 MW; 6AEAF6471BBB6FB CRC64;
Query Match	71.0%; Score 44; DB 16; Length 646;
Best Local Similarity	77.8%; Pred.No. 76;
Matches	7; Conservative 1; Mismatches 1; Indels 0; Gaps
QY	1 GGHPDPRPG 9



```

Db 239 GGHPRPPRG 247
|||||:|
SQ SEQUENCE 637 AA; 68945 MW; A52FEBB76F08D7C CRC64;
Query Match 70.2%; Score 43.5; DB 4; Length 637;
Best Local Similarity 80.0%; Pred. No. 90;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

RESULT 9
Q9BSV4 PRELIMINARY; PRT; 383 AA.
ID Q9BSV4 PRELIMINARY; PRT; 699 AA.
AC Q9BSV4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE PTB-associated splicing factor (Fragment).
GN SFPQ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=20465090; PubMed=11008015;
RA Shav-Tal Y., Lee B., Bar-Haim S., Vandekerckhove J., Zipori D.;
RT "Enhanced proteolysis of pre-mRNA splicing factors in myeloid cells.";
RL Exp. Hematol. 28:1029-1038(2000).
DR EMBL; AF272847; AAG17365.1; -.
DR HSSP; P11940; 1CVJ.
DR MGD; MGI:1918764; Sfpq.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 383
SQ SEQUENCE 383 AA; 43837 MW; 45E7CD226D9F841C CRC64;

Query Match 70.2%; Score 43.5; DB 11; Length 383;
Best Local Similarity 80.0%; Pred. No. 57;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 1 GGHPRPP-RG 9
|||||:|
Db 23 GGHFKPPHRG 32
|||||:|

RESULT 10
Q9BSV4 PRELIMINARY; PRT; 637 AA.
ID Q9BSV4 PRELIMINARY; PRT; 428 AA.
AC Q9BSV4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to splicing factor proline/glutamine rich (Polypyrimidine
tract-binding protein-associated) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004534; AA040434.1; -.
DR HSSP; P11940; 1CVJ.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 1

SQ SEQUENCE 699 AA; 75442 MW; 714F786264C63AA0 CRC64;
Query Match 70.2%; Score 43.5; DB 11; Length 699;
Best Local Similarity 80.0%; Pred. No. 98;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 1 GGHPRPP-RG 9
|||||:|
Db 220 GGHFKPPHRG 229
|||||:|

RESULT 12
Q9BSV4 PRELIMINARY; PRT; 428 AA.
ID Q9BSV4 PRELIMINARY; PRT; 428 AA.
AC Q9BSV4;
DT 01-JUN-2001 (TREMBLrel. 21, Created)
DT 01-JUN-2001 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to splicing factor proline/glutamine rich (Polypyrimidine
tract-binding protein-associated) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004534; AA040434.1; -.
DR HSSP; P11940; 1CVJ.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
FT NON_TER 1
FT NON_TER 1

```

RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Celnikier S.,  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY070499; AAL47970.1; -.  
DR FlyBase; FBgn0002643; mam.  
DR GO; GO:0005634; C:nucleus; IDA.  
DR GO; GO:0007500; P:mesoderm cell fate determination; IMP.  
DR GO; GO:0007399; P:neurogenesis; NAS.  
SQ SEQUENCE 428 AA; 43642 MW; D175E0DAEE9C6C9C CRC64;

Query Match 69.4%; Score 43; DB 5; Length 428;

Best Local Similarity 77.8%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGHPRPPRG 9

Db 407 GGFPFPHG 415

RESULT 13

Q65821 PRELIMINARY; PRT; 508 AA.  
ID Q65821;  
AC Q65821;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE UL44.  
GN UL44.  
OS Bovine herpesvirus 1.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=10320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COOPER;  
RX MEDLINE=93172350; PubMed=8382298;  
RA Fraefel C., Wirth U.V., Vogt B., Schwyzer M.,  
RT "Immediate-early transcription over covalently joined genome ends of  
RT bovine herpesvirus 1: the circ gene."  
RL J. Virol. 67:1328-1333(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JURA;  
RX MEDLINE=97164286; PubMed=9010999;  
RA Schwyzer M., Styger D., Vogt B., Lowery D.E., Simard C.,  
RA LaBoissiere S., Misra V., Vleck C., Paces V.;  
RT "Gene contents in a 31-kb segment at the left genome end of bovine  
RT herpesvirus-1."  
RL Vet. Microbiol. 53:67-77(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JURA;  
RA Schwyzer M., Paces V., Letchworth G.J., Misra V., Buhk H.J.,  
RA Lowery D.E., Simard C., Bello L.J., Thiry E., Vleck C.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JURA;  
RA Schwyzer M.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z54206; CAA90924.1; -.  
DR EMBL; AJ004801; CAA06089.1; -.  
DR InterPro; IPR001038; Gp13\_EHV.  
DR InterPro; IPR007110; IG-1-like.  
DR InterPro; IPR001654; Marek\_A.  
DR Pfam; PF02124; Marek\_A; 1.  
DR PRINTS; PR00666; IGLCPROTEIN.  
DR PROSITE; PS50835; IG\_LIKE; 1.  
SQ SEQUENCE 508 AA; 54119 MW; 76SD1B5DADE9F99D CRC64;

Query Match 69.4%; Score 43; DB 12; Length 508;

Best Local Similarity 87.5%; Pred. No. 88;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GGHPRPPR 8  
Db 121 GGRPRPPR 128

RESULT 14

O96545 PRELIMINARY; PRT; 537 AA.  
ID O96545;  
AC O96545;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative gag-related protein.  
OS Lymantria dispar (Gypsy moth).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;  
OC Lymantriidae; Lymantria.  
OX NCBI\_TaxID=13123;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=New Jersey Standard strain;  
RX MEDLINE=99306775; PubMed=10380107;  
RA Garner K.J., Slavicek J.M.;  
RT "Identification of a non-LTR retrotransposon from the gypsy moth."  
RL Insect Mol. Biol. 8:231-242(1999).  
DR EMBL; AF081103; AAC72920.1; -.  
DR InterPro; IPR006579; Pre\_C2HC.  
DR SMART; SMC0596; PRE\_C2HC; 1.  
SQ SEQUENCE 537 AA; 58036 MW; 3AF6C14C52023447 CRC64;

Query Match 59.4%; Score 43; DB 5; Length 537;

Best Local Similarity 66.7%; Pred. No. 92;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGHPRPPRG 9

Db 454 GGNPKPPQG 462

RESULT 15

O05089 PRELIMINARY; PRT; 545 AA.  
ID O05089;  
AC O05089;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Nocardioidea simplex (Arthrobacter simplex).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Propionibacteriaceae; Nocardioideae; Pimeleobacter.  
OX NCBI\_TaxID=2045;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IFO12069;  
RX MEDLINE=95319331; PubMed=7596291;  
RA Molnar I., Choi K., Yamashita M., Murooka Y.;  
RT "Molecular cloning, expression in Streptomyces lividans, and analysis  
RT of a gene cluster from Arthrobacter simplex encoding 3-  
RT ketosteroid-DELTA11-dehydrogenase, 3-ketosteroid-DELTA5-isomerase  
RT and a hypothetical regulatory protein."  
RL Mol. Microbiol. 15:895-905(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IFO12069;  
RA Dziadek J., Yamashita M., Murooka Y.;  
RT "Cloning, sequencing and characterization of the downstream region of  
RT KsdDI operon of Arthrobacter simplex."  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z93338; CAB07541.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 545 AA; 57318 MW; 55F441EBD5E3A19E CRC64;

Query Match 69.4%; Score 43; DB 2; Length 545;  
Best Local Similarity 87.5%; Pred. No. 94;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 HPRPPRGR 10  
|| |||||  
Db 456 HPPPPRGR 463

Search completed: April 6, 2004, 16:14:41  
Job time : 37.2451 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 / Search time 55.9813 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-7

Perfect score: 62

Sequence: 1 GGHPRPRGR 10

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003Bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	10	4	AAB48777 Human sal
2	62	100.0	17	4	AAB48783 Human sal
3	62	100.0	132	4	AB388848 Peptide #
4	62	100.0	132	4	AA32323 Peptide #
5	62	100.0	132	4	AA32323 Peptide #
6	62	100.0	132	4	AA32323 Peptide #
7	62	100.0	132	4	AA32323 Peptide #
8	62	100.0	132	4	AA32323 Peptide #
9	62	100.0	132	4	AA32323 Peptide #
10	62	100.0	132	4	AA32323 Peptide #
11	62	100.0	132	4	AA32323 Peptide #
12	62	100.0	132	4	AA32323 Peptide #
13	62	100.0	132	4	AA32323 Peptide #
14	62	100.0	132	4	AA32323 Peptide #
15	62	100.0	132	4	AA32323 Peptide #
16	62	100.0	132	4	AA32323 Peptide #
17	62	100.0	132	4	AA32323 Peptide #
18	62	100.0	132	4	AA32323 Peptide #
19	62	100.0	132	4	AA32323 Peptide #
20	62	100.0	132	4	AA32323 Peptide #
21	62	100.0	132	4	AA32323 Peptide #
22	62	100.0	132	4	AA32323 Peptide #
23	62	100.0	132	4	AA32323 Peptide #
24	62	100.0	132	4	AA32323 Peptide #
25	62	100.0	132	4	AA32323 Peptide #

26 45 72.6 30 3 AAY65870 Germline  
27 44 71.0 51 3 AAG58448 Zea mays  
28 44 71.0 74 4 AAM23839 Human EST  
29 44 71.0 74 4 AAM23850 Human EST  
30 44 71.0 74 4 AAM24011 Human EST  
31 44 71.0 121 4 AAU49746 Propionib  
32 44 71.0 121 6 ABM46265 Propionib  
33 43.5 70.2 355 4 AAM40927 Human pol  
34 43.5 70.2 707 4 AAM39141 Human pol  
35 43.5 70.2 707 7 ADE54941 Human Pro  
36 43.5 70.2 707 7 ADE54938 Human Pro  
37 43.5 70.2 707 7 ADE54944 Human Pro  
38 43.5 70.2 707 7 ADE54947 Human Pro  
39 43.5 70.2 707 7 ADE54935 Human Pro  
40 43.5 70.2 707 7 ADE54950 Human Pro  
41 43.5 70.2 707 8 ADE77113 Human pro  
42 43 69.4 112 4 AAU57069 Propionib  
43 43 69.4 112 6 ABM53588 Propionib  
44 43 69.4 202 2 AAR21441 Construct  
45 43 69.4 255 4 AAU47216 Propionib

## ALIGNMENTS

RESULT 1  
AAB48777  
ID AAB48777 standard; peptide; 10 AA.  
XX AAB48777;  
DT 09-MAR-2001 (first entry)  
DE Human saliva PRP-1 fragment (residues 99-108), SEQ ID NO:7.  
XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX Homo sapiens.  
XX WO200069890-A1.  
PD 23-NOV-2000.  
XX 11-MAY-2000; 2000WO-SE000930.  
XX 17-MAY-1999; 99SE-00001773.  
XX (STRO/) STROEMBERG N.  
XX (JOHA/) JOHANSSON I.  
XX Stroemberg N, Johansson I;  
XX WPI; 2001-031923/04.  
XX New oligopeptides comprising 2 arginine residues from degradation of  
XX proline-rich proteins, useful for preventing dental caries.  
XX Claim 4; Page 24; 36pp; English.  
XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
XX AAB48783) which contain at least two arginine residues and which protect  
XX against dental caries. PRPs (proline-rich proteins) are salivary proteins  
XX encoded by six clustered genes on chromosome 12p13.2 and are potential  
XX determinants of a person's susceptibility to dental caries. PRPs are  
XX degraded by Actinomyces and Streptococcus species to small peptide  
XX fragments. These are metabolised by oral bacteria for nutritional  
XX purposes, with certain bacterial species generating ammonia via the  
XX catabolism of arginine. The peptides of the invention, being arginine-  
XX rich, can also be converted to ammonia by these bacteria. The ammonia  
XX thus formed raises the pH at the dental surface, thereby protecting the  
XX teeth against caries. Sequences AAB48771-948783 represent the PRP-1-